Cuaderno 55

HILARY PUTNAM

EL SIGNIFICADO Y LAS CIENCIAS MORALES

Traducción de
ANA ISABEL STELLINO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MÉXICO 1991

EL SIGNIFICADO Y LAS CIENCIAS MORALES

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

Colección: CUADERNOS

Director: DR. LEÓN OLIVÉ

Secretario: DR. AMBROSIO VELASCO

Cuaderno 55

HILARY PUTNAM

EL SIGNIFICADO Y LAS CIENCIAS MORALES

Traducción de
ANA ISABEL STELLINO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MÉXICO 1991

Título original:
Meaning and the Moral Sciences
Routledge & Kegan Paul, Londres, 1978
© Hilary Putnam, 1978

DR © 1991 Universidad Nacional Autónoma de México
Circuito Mario de la Cueva
Ciudad de la Investigación en Humanidades
Ciudad Universitaria, 04510, México, D. F.
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS
Impreso y hecho en México
ISBN 968-36-2250-X

PREFACIO

Las conferencias que conforman este volumen fueron escritas en un periodo de dos años extraordinariamente memorables para mí. La invitación a dictar las Conferencias John Locke en la Universidad de Oxford en 1976, combinada con el afortunado accidente de hallarme en mi periodo sabático otorgado por la Universidad de Harvard, me dieron tiempo y estímulo para escribir durante un periodo prolongado. Las Conferencias Locke surgieron de reflexiones propiciadas por el motivador artículo de Hartry Field sobre la teoría de la verdad en Tarski. Una vez que bosquejé el primer borrador, me asombré de las conclusiones a las que había llegado (por ejemplo, que posee alguna relación con la tesis de la 'indeterminación de la traducción' de Quine) y me sentí complacido de poder conectar en las conferencias mi interés por la epistemología de la ética (que, debo agregar, se encuentra en un estadio preliminar y poco desarrollado) con mi desagrado por lo que llamaré el 'utopismo científico'. Aun así cuando llegué a Oxford para dictar seis conferencias durante el felizmente denominado Trimestre Hilary de 1976, sólo tenía los primeros bosquejos de cuatro o cinco. De una u otra manera logré presentarme con las conferencias preparadas, y sé que la paz, el estímulo y la calidez de la atmósfera de Oxford me ayudaron a trabajar con intensidad.

Tantas personas me brindaron su amistad y su gentileza durante ese corto periodo, que me es imposible mencionarlas a todas (y tampoco considero que una lista pudiera ser de interés general). Pero debo agregar que la oportunidad de dictar las Conferencias Shearman en la Universidad de Londres me dio la excusa para visitar a una buena cantidad de amigos en Londres y que este estímulo (tanto social como filosófico) fue importante para mí. Las conversaciones

que tuve con mis amigos en Cambridge me dejaron cavilando hasta muchos meses después.

También quisiera añadir que ser Profesor John Locke no fue sólo un honor, sino que para un anglófilo y fanático de Zuleika Dobson como yo, fue una espléndida oportunidad para realizar muchas fantasías eruditas. En particular, vivir en el *College* universitario resultó una experiencia que siempre estimaré (y *tengo* que mencionar a George Cawkwell, quien hizo mucho por reducir el choque cultural y para hacerme sentir como en casa). Ese invierno fue el de la sequía (aunque nosotros no nos percatamos de ello), así que no hubo lluvia, ni nieve, y la primavera llegó antes, a mediados de febrero. Hubo, entonces, maravillosas oportunidades para apreciar la arquitectura de Oxford, para pasar fines de semana en el campo y para otras actividades.

Después de Oxford viajé a Israel, país que no conocía, donde una vez más se presentó el riesgo de un choque cultural, felizmente sorteado con la ayuda de cálidos amigos (tengo que mencionar a Yehuda Elkana) y la belleza del campo y de Jerusalén. Dado que en ese momento la Universidad de Jerusalén auspició una conferencia internacional sobre filosofía del lenguaje (desafortunadamente, Yehoshua Bar Hillel murió a inicios de aquel año, lo que impidió su participación en ella), tuve ocasión de encontrarme otra vez con Peter Strawson y Michael Dummett.

Como no había podido asistir a la mayoría de las Conferencias William James impartidas por Dummett (que se realizaron en Harvard cuando yo estaba en Oxford), me sentí encantado de que viniera a Jerusalén y comentara "Referencia y entendimiento", que fue, en buena medida, una respuesta a las Conferencias William James (a las que oí y también a las transcripciones de las otras). Sus comentarios fueron muy importantes para mi pensamiento y, aunque todavía existen muchos puntos de desacuerdo en nuestras perspectivas, el giro representado por "Realismo y razón", es en parte el resultado de nuestro diálogo.

Sin embargo, "Realismo y razón" no fue escrito durante el año sabático, sino en el otoño de 1976. Y el acontecimiento más importante para mi trabajo durante ese periodo fue la sugerencia de Nelson Goodman de que él, Willard van Quine y yo organizáramos una serie de encuentros en los que pudiéramos discutir las posturas de unos y otros. Sé que estos tres "espectáculos" (como irreverentemente los llamaron los estudiantes) representaron un tremendo

estímulo, y que esta intensa interacción con la visión de Goodman (y con la de Quine, aunque la influencia de su pensamiento en mi trabajo es de viejo cuño) también delineó "Realismo y razón".

Confío en haber transmitido la impresión de que disfruté la redacción de estas conferencias. Espero que el lector disfrute su lectura.

Y, por supuesto, mi gratitud a la Fundación Nacional para las Humanidades por haber permitido que esto fuera posible.

INTRODUCCIÓN

Antes de Kant casi todos los filósofos suscribían a la concepción de que la verdad es algún tipo de correspondencia entre las ideas y 'lo que es el caso'. No importa qué tan misteriosa pueda parecer la *naturaleza* de la 'correspondencia', la naturalidad de la idea es innegable. Existe un mundo en el exterior y lo que decimos o pensamos es 'verdadero' cuando captamos *la forma en que es* y 'falso' cuando no corresponde a *la forma en que es*.

Con Kant surge una nueva perspectiva, la de que la verdad es radicalmente dependiente de la mente. No es que, desde su punto de vista, la mente pensante *componga* al mundo, pero tampoco se limita a reflejarlo.

En la actualidad, las concepciones sobre la verdad pueden dividirse en dos clases: la 'realista', que la interpreta como un tipo de correspondencia con lo que es el caso, y la 'verificacionista' que la interpreta como, por ejemplo, lo que podría *verificarse* bajo condiciones ideales de investigación. (Elijo la formulación de Peirce como ejemplo de perspectiva verificacionista porque me parece la más defendible, pero existen, por supuesto, muchas versiones tanto del punto de vista realista como del verificacionista en la literatura filosófica.)

Sin embargo, en las primeras décadas del siglo XX hubo muchos filósofos de corte empirista que no se adscribían a ninguna de estas concepciones. Estos filósofos *rechazaban* la noción total de 'verdad' como 'metafísica'. (Un ejemplo es John Dewey, quien siempre habló de 'aseverabilidad justificada', y no de 'verdad'.) En particular, muchos de los positivistas del Círculo de Viena compartieron esta actitud de recelosa hostilidad hacia el concepto de verdad hasta conocer el trabajo de Alfred Tarski.

El lector encontrará una exposición del trabajo de Tarski en la primera conferencia de este volumen; lo importante es que este trabajo fue tomado por los filósofos como *deflacionario*. Lo que Tarski hace es mostrar cómo, en el contexto de un lenguaje formalizado, se puede definir 'verdadero' (o un predicado que pueda usarse en su lugar) aplicando *únicamente* las nociones del lenguaje-objeto y de las matemáticas puras. En particular, Tarski no toma ninguna noción semántica —por ejemplo: 'designa', 'representa', o 'refiere a'— como primitiva (aunque 'refiere a' queda *definida*, en términos de las nociones no semánticas, en el curso de su trabajo). Así, quien acepte las nociones de cualquier lenguaje objeto de que se trate, y éste podría elegirse arbitrariamente, puede también entender 'verdadero' como lo definió Tarski para ese lenguaje objeto. 'Verdadero' posee la misma legitimidad que cualquier otra noción de ciencia de primer orden.

Lo esencial es que Tarski verifica la exactitud de su 'definición de verdad' para cualquier caso particular mediante la constatación de que satisface cierta condición de equivalencia. Esta condición consiste en decir de cualquier oración que, para que sea verdadera, debe ser equivalente (de hecho, Tarski requiere que sea *demostrablemente* equivalente) a la oración misma. Para usar su famoso ejemplo, si 'La nieve es blanca' es una oración del lenguaje objeto, entonces 'verdadero' podrá ser definido de tal forma (para ese lenguaje objeto) que, en el lenguaje en que se da la definición y al que Tarski llama *metalenguaje*, se vuelva *demostrable* que

'La nieve es blanca' es verdadera si y sólo si la nieve es blanca.

Un ejemplo del uso que los filósofos empiristas hacen de esto es el siguiente: un pragmatista menor llamado Felix Kaufmann ha argumentado que 'verdadero' es una noción metafísica (entendiendo este término en sentido peyorativo), sobre la base de que (según afirma) las adscripciones al predicado 'verdadero' nunca podrán ser verificadas o falseadas. Kaufmann sostiene que podemos decir si un juicio está *confirmado* hasta cierto punto, pero dado que una confirmación *perfecta* resulta imposible, jamás podremos verificar si el juicio es absolutamente *verdadero*. Apoyado en el trabajo de Tarski, Carnap replicó a Kaufmann como sigue: los empiristas liberales (incluyendo al propio Kaufmann) no requieren que las adscripciones de un predicado sean verificables con certeza a fin de que el predicado

sea admitido en un lenguaje empírico. Por ejemplo, podemos demostrar con una probabilidad bastante alta para cualquier propósito práctico (y aun científico) si la corriente fluye o no por un cable; así el predicado 'es tal la corriente que fluye', en tanto predicado que se aplica a los cables de cobre, es aceptado como empíricamente significativo. Pero si

'La corriente fluye a través de este cable'

tiene una probabilidad, digamos, de 0.999, entonces la oración del metalenguaje

'La corriente fluye a través de este cable' es verdadera

tiene exactamente la misma probabilidad de 0.999 por la condición de equivalencia. Kaufmann tiene razón al decir que no es *verificable con certeza* que la oración " 'La corriente fluye a través de este cable' es verdadera" sea correcta, pero es *verificable con probabilidad*, a saber, mediante la comprobación de la oración correspondiente del lenguaje objeto 'La corriente fluye a través de este cable' y usando la condición de equivalencia.

Y así sucedió que en los años treinta, Carnap, Popper, Hempel, etc., aceptaron la noción de 'verdadero' como una noción que aun los más escrupulosos empiristas podían utilizar felizmente. Pero la aceptaron *porque pensaban que no desempeñaba ninguna función filosófica*. De hecho, si uno identifica la teoría de la verdad por correspondencia con la condición de equivalencia, entonces la primera tampoco desempeña ninguna función filosófica. Que *'La nieve es blanca' es verdadera si y sólo si la nieve es blanca* no es más que una tautología. Ésta no nos dice nada acerca de cómo se *usa* o *entiende* 'La nieve es blanca', nada acerca de su aseverabilidad, etcétera.

No resulta sorprendente que pronto se produjera una reacción realista contra la pretensión de que la teoría de la verdad carece de interés filosófico (y que la misma noción de verdad es filosóficamente neutral). En un documento que presenté ante el club de estudiantes de filosofía de la Universidad de Oxford en 1960 y que titulé '¿Corresponden las afirmaciones verdaderas a la realidad?', escribí:

> Quiero explorar el problema de si tiene algún sentido el punto de vista tradicional de que una afirmación es verdadera cuando se corresponde con la realidad. Dos opiniones parecen ser las más difundidas: a) que

se puede encontrar algún sentido en esta perspectiva y que la 'concepción semántica de la verdad' de Tarski *extrae* de ella algún sentido (tanto como pudiera esperarse); b) que esa concepción se derrumba tan pronto como uno formula preguntas escudriñadoras sobre la naturaleza de la supuesta 'correspondencia'. Trataré de mostrar que ambas opiniones son incorrectas.

La estrategia que seguí, y que han seguido desde entonces algunos filósofos, consiste en considerar al hablante de un lenguaje como si construyera una representación simbólica de su medio. Aquél no podría proceder así a menos que interactúe causalmente con el segundo, y la precisión e imprecisión de su representación afectará la viabilidad y el éxito de su actividad cuando trate con su ambiente; por lo tanto, esta explicación de la relación de los hablantes de un lenguaje con el mundo forma parte de un modelo causal de la conducta humana. En la medida en que la correspondencia presupuesta entre las representaciones mentales del hablante y sus referentes externos forma parte del modelo, el realismo se vuelve entonces una *hipótesis empírica*. En 'Referencia y entendimiento', un ensayo incluido en este volumen, intenté explicar con claridad este tipo de concepción y compararla con otras. En la segunda de mis conferencias Locke de este volumen, el lector también encontrará una discusión de la idea de que el realismo es una hipótesis empírica y del trabajo de Richard Boyd, quien ha propuesto una perspectiva afín.

Debo señalar que la estrategia no es mostrar que Tarski está *equivocado*. Más bien, argumento que este autor nos ha dado una explicación completamente correcta de la lógica formal del concepto 'verdadero'. Pero la lógica formal del concepto no es todo lo que hay en la noción de verdad. El trabajo de Tarski requiere complementarse filosóficamente y cuando proporcionamos esta complementación, vemos que la noción de verdad no es filosóficamente neutral, y que se necesita una explicación de la correspondencia para entender cómo opera el lenguaje y cómo lo hace la ciencia.

A inicios de la década de 1970 apareció un artículo extraordinariamente brillante y provocativo de Hartry Field, 'La teoría de la verdad de Tarski', que contenía un serio ataque al trabajo de éste desde la perspectiva realista. Field afirmaba que la parte del trabajo de Tarski que precisamente explota la condición de equivalencia es insatisfactoria y que la misma debe ser remplazada por un tipo di-

ferente de teoría si queremos establecer en la teoría del lenguaje patrones que resulten respetables a nivel científico.

Aunque no estoy de acuerdo con la conclusión de Field, considero que este artículo representa un profundo tema de debate, y mis conferencias Locke fueron concebidas en su origen justamente como un examen detallado del problema planteado por Field, el de tratar de encontrar (o pedir que busquemos) una teoría 'fisicalista' de la referencia, es decir, una teoría para la cual el hablante sólo se refiere a una cosa al utilizar un término si el uso que le da se encuentra en una relación causal definida (por la ciencia empírica, en el mismo espíritu en que ésta ha 'definido' *agua* como H_2O y *temperatura* como energía cinética molecular) con ese algo, o con cosas de la misma clase que ese algo (una 'misma clase' cuyo sentido también tendría que haber sido definido). Lo que yo no había previsto fue que mientras discutía la crítica de Field a Tarski (y explicaba mis razones para refutarla), me vería arrastrado al debate del cientificismo de las ciencias sociales, según se halla representado en la celebrada máxima de Mill de que 'El estado de atraso de las ciencias morales' sólo puede ser remediado si se imita el método de la física. El lector se percatará de que las conferencias Locke se mueven gradualmente de una discusión específica sobre las nociones de verdad y referencia, a cuestiones relacionadas con el carácter de la investigación lingüística en general e incluso de la investigación científico social en general.

He incluido en este volumen un ensayo sobre temas morales y literarios, a pesar de las insuficiencias que contiene, porque quiero hacer explícita la conexión que, a mi entender, existe entre la errónea concepción que pienso que tenemos acerca de las ciencias sociales y la concepción, también equivocada —la cual pertenece a un fenómeno cultural mayor—, conocida como la 'distinción hecho/valor'. La conexión reside en la presunción de que las palabras 'verdad' y 'conocimiento' son de algún modo *propiedad* de la 'ciencia' (construida a partir de un modelo particular). Pienso que Aristóteles estaba inequívocamente en lo cierto al sostener que la ética tiene que ver con la forma de vivir y con la felicidad humana y también al afirmar que este tipo de *conocimiento* (conocimiento práctico) es diferente del teórico. Una visión que acepte que la esfera del conocimiento es más amplia que la esfera de la 'ciencia' me parece una necesidad cultural si queremos llegar a una postura sensata y humana de nosostros mismos *o* de la ciencia.

El ensayo final en este volumen representa un giro reciente de mi pensamiento. Dicho burdamente, la pretensión es que una teoría 'verificacionista' de la verdad y una teoría de la verdad por correspondencia (en el sentido del realismo empírico) *no* son incompatibles. Considerando que Kant asumió una postura que abarcaba a la concepción de la correspondencia de la verdad *dentro* del reino empírico (al menos, según mi interpretación), *y* una acentuación del supuesto de que toda verdad depende de la mente, la postura con la que concluyo posee (quiero creer) algo de kantiano. Pero se trata de un kantismo desmitificado, sin 'cosas en sí' ni 'yoes trascendentales'.

Estos ensayos cubren, obviamente, diversos temas. Pero exceptuando 'Ciencia, literatura y reflexión', todos fueron escritos en un periodo de dos años y están íntimamente relacionados. Lo cierto es que el gran problema metafísico del realismo, la naturaleza del lenguaje y de la referencia, y el carácter del conocimiento sobre nosotros mismos y nuestro saber, *son* temas que se superponen, interrelacionan y configuran entre sí. Espero que el lector termine este libro convencido de que así es.

PRIMERA PARTE

SIGNIFICADO Y CONOCIMIENTO
(Conferencias John Locke, 1976)

PRIMERA CONFERENCIA

La naturaleza de la verdad es un problema muy antiguo en la filosofía, pero no fue sino hasta el presente siglo cuando los filósofos y los lógicos intentaron separarlo de los problemas de la naturaleza del conocimiento y de la naturaleza de la creencia garantizada. A un filósofo del siglo XIX le hubiera parecido ridículo que uno pudiera tener una teoría de la verdad que fuera neutral con respecto a las cuestiones epistemológicas, e incluso con respecto al gran problema metafísico del realismo contra el idealismo. Pero esto es, precisamente, lo que la más prestigiosa teoría de la verdad, la teoría de Tarski,[1] pretende sostener.

Aunque para presentar esta teoría con propiedad se necesita un cierto grado de lógica complicada, resulta fácil explicar una de sus ideas principales, la idea de 'desentrecomillado'. Tomemos cualquier oración, digamos *La nieve es blanca*. Entrecomillemos la oración como sigue:

'La nieve es blanca'.

Ahora agreguemos las palabras 'es verdadera', así:

'La nieve es blanca' es verdadera.

La oración resultante es ella misma verdadera si y sólo si la oración original lo es. Más aún: es *aseverable* si y sólo si la oración original es aseverable; es probable en un grado *r* si y sólo si la oración original es probable en un grado *r*, etc. De acuerdo con Tarski, Carnap, Quine, Ayer y teóricos semejantes, saber estos hechos constituye la

[1] *Cfr.* Tarski, "The Concept of Truth in Formalized Languages", reimpreso en su *Logic, Semantics, Metamathematics*, Oxford University Press, 1956, pp. 152–278.

clave para entender las palabras 'es verdadero'. En suma, para entender *P es verdadero*, donde *P* es una oración entrecomillada, sólo hay que 'desentrecomillar' *P* —eliminar las comillas (y borrar 'es verdadero').

Por ejemplo, ¿qué significa

'La nieve es blanca' es verdadera?

Significa

La nieve es blanca

¿Qué significa

'Existe un mundo externo real' es verdadera?

Significa

Existe un mundo externo real

Etcétera.

La proposición sugerida por los teóricos del 'desentrecomillado' en este caso consiste en que una respuesta a la pregunta '*¿qué significa decir* que algo es verdadero?' no necesita hallarse referida a una concepción sobre lo que a su vez ese algo significa ni sobre cómo habrá o no de verificarse. Se puede dar una interpretación materialista de 'La nieve es blanca; se puede creer que 'La nieve es blanca' es verificable o que resulta únicamente falsificable pero no verificable; o que sólo puede confirmarse en un grado entre cero y uno; o ninguna de las opciones precedentes, pero 'La nieve es blanca es verdadera' sigue siendo equi-aseverable a "'La nieve es blanca' es verdadera". Desde esta perspectiva, 'verdadero' constituye, para nuestra sorpresa, una noción *filosóficamente neutral*. Se trata sólo de un recurso para el 'ascenso semántico' —para 'elevar' afirmaciones del 'lenguaje objeto' al 'meta lenguaje', y el recurso no nos compromete epistemológica ni metafísicamente.

Ahora bosquejaré la segunda idea principal de la teoría de Tarski. En ésta, 'verdadero' es un predicado de oraciones y, para precisar la teoría, esas oraciones tienen que estar formuladas en un lenguaje formalizado *L*. (Cómo aplicar la teoría a los lenguajes naturales representa hoy un gran tema de debate entre filósofos y lingüistas.) Ahora bien, un 'lenguaje' en este sentido posee un número *finito* de predicados no definidos o 'primitivos'. Para simplificar, supongamos

que nuestro lenguaje *L* sólo tiene dos predicados primitivos: 'es la luna' y 'es azul'. Para predicados *P*, la locución

P se refiere a *x*

cuya íntima conexión con la palabra 'verdadero' puede ponerse de manifiesto mediante la frase 'es verdadero de' en lugar de 'se refiere a', de modo que

P es verdadero de *x*

puede también explicarse empleando la idea del desentrecomillado: si *P* es el predicado 'es la luna', tenemos:

'Es la luna' se refiere a *x* si y sólo si *x* es la luna.

Y si *P* es el predicado 'es azul', tenemos:

'Es azul' se refiere a *x* si y sólo si *x* es azul.

Así la predicación 'metalingüística'

'Es la luna' se refiere a *x*

resulta equivalente a la predicación del 'lenguaje objeto'

x es la luna.

Digamos que *P se refiere primitivamente* a *x* si *P* es un predicado primitivo (en el caso de nuestro lenguaje *L*, 'es la luna' o 'es azul') y *P* se refiere a *x*. Entonces una *referencia primitiva* puede ser definida *para nuestro ejemplo particular L* mediante la presentación de una lista:

Definición:

P se refiere primitivamente a x si y sólo si (1) *P* es la frase 'es la luna' y *x* es la luna, o (2) *P* es la frase 'es azul' y *x* es azul.

Y para cualquier *lenguaje formalizado particular* puede darse una definición semejante de la referencia primitiva, una vez que se nos haya proporcionado una lista de los predicados primitivos de ese lenguaje.

Para explicar con propiedad el resto de la idea de Tarski se requiere de la lógica y las matemáticas. Aquí lo expondré a *muy* grandes rasgos.

Los predicados no primitivos de un lenguaje se construyen a partir de los primitivos a través de diversos recursos: las funciones de

verdad y los cuantificadores. Supongamos, por dar un ejemplo, que los únicos recursos son la disyunción y la negación: formamos así los predicados '*P o Q*', y '*no P*' del predicado *P*. Luego definimos la referencia como sigue:

(I) Si *P* no contiene ninguna conectiva lógica, *P se refiere a x* si y sólo si *P* está referido primitivamente a *x*.

(II) *P o Q se refiere a x* si *P* se refiere a *x* o *Q* se refiere a *x*.

(III) *No P se refiere a x* si *P* no está referido a *x*.

Cuando se convierte esta definición inductiva[2] en una definición propiamente dicha aparece gran parte de la lógica técnica; baste indicar que esto puede hacerse. El resultado es una definición de la 'referencia' *para un lenguaje particular* —una *definición que no utiliza términos semánticos* (ningún término de la misma familia a la que pertenece 'verdadero' o ' se refiere').

Finalmente, suponiendo que nuestro lenguaje simple lo fuera tanto[3] que la totalidad de las oraciones se presentaran bajo la forma: *para toda x, Px, para algunas x, Px*, o bajo las funciones de verdad de éstas (donde *P* es un predicado), entonces *verdadero* podría definirse como sigue (desde luego, Tarski consideraba en realidad lenguajes mucho más ricos):

(I) *para toda x, Px* es *verdadera* si y sólo si, para toda *x*, *P* se refiere a *x*.

[2] En lógica se llama 'explícita' a una definición si suministra una regla para *eliminar* la expresión definida —remplazándola con vocabulario primitivo— en todos los contextos. Las definiciones llamadas 'inductivas' no son realmente definiciones en este sentido, sino más bien especificaciones de la extensión que tiene una noción por inducción matemática —en este caso, una inducción del número de conectivas lógicas en el predicado. Pero tales definiciones inductivas pueden ser *convertidas* en definiciones explícitas mediante recursos más o menos complicados de la lógica matemática.

[3] Este lenguaje se llama *monádico* (porque únicamente aparecen predicados de una sola situación) y *uniforme* (porque ningún cuantificador se encuentra dentro del radio de otro cuantificador). La última restricción no es esencial porque cualquier fórmula monádica es equivalente a una fórmula monádica *uniforme* (para una comprobación de esto véase W. V. Quine, *Methods of Logic*, tercera ed., Nueva York, Holt, 1967, parte II); pero circunscribirse a fórmulas *monádicas* sólo es esencial para eludir todas las complicaciones de la teoría de Tarski (el uso de secuencias infinitas, etc.), y quiero evitarla aquí por razones de exposición.

(II) *para algunas x*, *Px* es *verdadera* si y sólo si, para alguna *x*, *P* se refiere a *x*.

(III) si *p* y *q* son oraciones, *p* o *q* es *verdadera* si *p* es verdadera o *q* es verdadera; y *no p* es verdadera si *p* no es verdadera.

Aunque he omitido las matemáticas incluidas en el trabajo de Tarski (de qué manera se transforma una 'definición inductiva' como la anterior en una definición 'explícita' del tipo 'algo es verdadero si y sólo si...', donde 'verdadero' y 'se refiere' no aparecen en '...') y he pasado por alto las inmensas complicaciones que surgen cuando el lenguaje tiene *relaciones* —predicados de dos situaciones (o de tres situaciones, etc.)—, espero haber transmitido tres ideas:

(1) 'Verdad' y 'referencia' se definen *para un lenguaje particular por vez*. No definimos la relación 'verdadero en L' para un L *arbitrario*.

(2) La referencia primitiva se define 'por una lista'; y en general, referencia y verdad son definidas por inducción a partir del número de conectivas lógicas que existen en el predicado u oración, empezando con la referencia primitiva.

(3) La definición 'inductiva' realizada mediante un sistema de cláusulas como (I), (II), (III), (I'), (II'), (III'), puede convertirse en una 'definición explícita' *bona fide* utilizando los recursos técnicos de la lógica.

Como prueba de que lo efectuado es correcto, resulta fácil derivar de la definición de 'verdadero' el siguiente teorema:

> *'Para algunas x, x es la luna'* es *verdadero* si y sólo si, para algunas *x*, *x* es la luna.

Y en efecto, se puede derivar de la definición de verdadero que

(*T*) '*P*' es verdadero si y sólo si *P*

cuando se sustituye la letra nominal '*P*' por cualquier oración de nuestro lenguaje L.

El hecho de que esto deba ser así —que el esquema anterior (*T*) sea tal que todas sus instancias sean una consecuencia de la definición de 'verdad'— constituye el 'criterio de adecuación' de Tarski (el famoso 'criterio *T*') para las definiciones de 'es verdadero'.

Observemos que aunque la idea del desentrecomillado podría chocarnos al principio por su trivialidad, la teoría de Tarski no es en modo alguno superficial. La razón reside en que esa idea se limita a indicarnos que el criterio *T* es correcto; pero no nos explica cómo definir 'verdadero' de modo que se satisfaga el criterio *T*. Por sí mismo el desentrecomillado tampoco nos permite eliminar el término 'verdadero' de todos los contextos en que aparezca. *La nieve es blanca' es verdadera* equivale a *La nieve es blanca*; pero ¿a qué oración que *no contenga la palabra 'verdadero'* (o cualquier otro término semántico) es equivalente el siguiente enunciado?: *Si las premisas en una inferencia de la forma p o q, no-p/∴ q son ambas verdaderas en L, la conclusión es también verdadera en L*. El método de Tarski nos da un equivalente para *esta* oración, y para otras en las cuales 'es verdadero' aparece con variables y cuantificadores, y eso es lo que el desentrecomillado por sí mismo no hace.

Hay muchos problemas en la teoría de Tarski que me han molestado durante años. Uno de ellos, que *no* me parece demasiado serio, consiste en que 'verdadero' se toma como un predicado de *oraciones* (es decir, cadenas de signos escritos); hablando estrictamente, lo que se analiza no es 'es verdadero' como un predicado de enunciados, sino 'expresa un enunciado verdadero'. Pero asumo que, aunque va en contra de la costumbre calificar a las oraciones como verdaderas o falsas, este uso resulta muy claro, además de que este pequeño alejamiento de la costumbre probablemente sea inevitable en *cualquier* reconstrucción (por ejemplo, el uso de la palabra 'enunciado').[4] Una objeción más importante se refiere a que la teoría no permite oraciones que no son *ni* verdaderas *ni* falsas (por ejemplo, 'El número de árboles en Canadá es par'), u oraciones que contienen palabras indicativas (como 'ahora', 'aquí', 'yo'). De hecho, la teoría sólo puede aplicarse en su forma canónica a lenguajes en los que todo predicado esté bien definido y no sea indicativo. Pero creo que se la puede modificar para aplicarla a lenguajes de otros tipos, aunque ésta no es la clase de problema que voy a abordar en estas conferencias. Tampoco señalaré las distinciones que existen entre la teoría original de Tarski y una variante reciente y muy prolija propuesta por Saul Kripke. La

[4] Para una exposición más completa de este punto véase mi *Philosophy of Logic*, Nueva York, Harper and Row, 1971, pp. 18–21.

crítica que quiero tratar en detalle y con la cual alguna vez estuve de acuerdo es la presentada por Hartry Field.[5]

Field reconoce que Tarski logró algo de importancia filosófica al mostrar cómo definir las nociones semánticas de referencia y verdad apegándose a la noción semántica de la referencia primitiva. Pero objeta que Tarski se equivocó al pensar que había aclarado filosóficamente la referencia primitiva. Éste es el parágrafo crucial del artículo de Field:[6]

> Ahora bien, habría sido fácil para un químico de fines del siglo pasado dar una 'definición de valencia' en la forma siguiente:
>
> (3) $(\forall E)(\forall n)(E$ tiene valencia $n = E$ es potasio y n es $+1$, o ... o E es azufre y n es $-2)$
>
> donde en los espacios va una lista de cláusulas similares, una para cada elemento. Pero aunque ésta es una definición de valencia correcta en términos de su extensón, no habría sido una reducción aceptable, y si se hubiera hallado que no existía otra posibilidad, es decir, si todos los esfuerzos para explicar la valencia de acuerdo con las propiedades estructurales de los átomos se hubieran revelado como fútiles, los científicos habrían tenido que decidirse eventualmente por (a) renunciar a la teoría de la valencia, o bien (b) remplazar la hipótesis del fisicalismo por otra (¿el quimicalismo?). Es parte de la metodología científica resistirse a optar por (a) en tanto la noción de valencia sirva a los propósitos para los que fue concebida (es decir, en tanto demuestre su utilidad para ayudarnos a caracterizar los compuestos químicos según sus valencias). Pero la metodología no consiste en oponerse a (a) y a (b) confeccionando listas como (3), sino buscar una verdadera reducción. Se ha comprobado que esta metodología es extraordinariamente fructífera para la ciencia y creo que la perderemos a menos que tomemos conciencia de que necesitamos añadir a T_1 y T_2 las teorías de la referencia primitiva si queremos establecer la noción de verdad como una noción fisicalista aceptable.

Field está objetando que la noción de *referencia primitiva* de Tarski (en nuestro ejemplo, P se refiere primitivamente a x si P es la frase 'es la luna' y x es la luna, o P es la frase 'es azul' y x es azul) resulta exactamente igual a la 'absurda definición de valencia' (3). En suma, Field afirma que tenemos *niveles bajos* en la teoría del lenguaje, y que

[5] *Cfr.* Field, 'Tarski's Theory of Truth', *Journal of Philosophy*, vol. 69, núm. 13, 1972, pp. 347–375.

[6] *Ibid.*, pp. 362–363.

deberíamos poseer los mismos niveles (fisicalistas) que existen en otras ciencias naturales, especialmente si, como buenos fisicalistas, consideramos al lenguaje como un fenómeno natural.

Encuentro esto fascinante porque ilumina una cuestión que había estado sumergida para la filosofía durante mucho tiempo y que se puso de manifiesto en el siglo XX. ¿La *referencia* sólo constituye una relación entre una cosa, que resulta ser una palabra (digamos, 'Mond') o el acontecimiento particular de que esa palabra sea emitida, y otra cosa (digamos, la luna)? Si es así, ¿es una relación que forma parte del orden natural causal en igual medida que la relación 'está químicamente ligado a'? ¿Debe estudiársela de la misma manera? Y si no es así, ¿estamos considerando al lenguaje como algo trascendental (ya sea en el sentido kantiano o en cualquier otro)? ¿O debemos dejar de hablar de 'referencia' por completo? Field opta por la concepción 'fisicalista' extrema que contesta 'sí' a mis tres primeras preguntas; aun si está equivocado, esto es lo que da relevancia e interés a su artículo.

Quiero considerar la ingeniosa réplica (inédita) de Stephen Leeds a Field. En efecto, aquél pregunta *para qué* queremos una noción de verdad. Según los propósitos de la lógica deductiva, por ejemplo, con el fin de decir 'en inferencias de la clase X, la conclusión es verdadera cualesquiera que sean las premisas'. Para expresar acuerdos, cuando no sabemos exactamente qué oración fue emitida ('lo que dijo era verdad'), etcétera.

Lo que sigue es una sagaz observación de Leeds: si tuviéramos un metalenguaje con *conjunciones infinitas* y *disyunciones infinitas* (infinitos contables), ¡no necesitaríamos el término 'verdadero'! Por ejemplo, si quisiéramos expresar 'lo que dijo era verdadero', lo haríamos así:

(1) [Él dijo 'P_1' y P_1] o [Él dijo 'P_2' y P_2] o ...

donde la disyunción en (1) contiene un disyunto para cada oración 'P_1' del lenguaje objeto.

Pero *no podemos* hablar en la práctica mediante disyunciones infinitas. Por ello buscamos en cambio una expresión finita equivalente a (1). Ahora bien,

(2) *Para algunas x él dijo x* y *x es verdadero*

será equivalente a (1) toda vez que para cada i (i = 1, 2, 3, ...)

(3) 'P_i' es verdadero si y sólo si P_i

resulte correcta. Pero éste es justamente el 'criterio *T*' de Tarski, el famoso criterio de adecuación. En suma, cualquier definición de 'verdadero' estará bien *siempre y cuando* se satisfaga el punto (3). Y no importa *cómo* se presente la definición de 'verdadero' —podría hallarse en términos de 'referencia primitiva' donde esta última fuera definida por una *lista* análoga a la 'absurda definición de valencia' de Field—, porque toda vez que *aceptemos* (3) (es decir, aceptemos las instancias completas de (3)), sabremos cómo expresar todas las disyunciones infinitas que no somos capaces de formular con las expresiones finitas que *podemos* decir (según el modelo de expresar (1) diciendo (2)).

El meollo de la réplica de Leeds a Field es éste: la valencia constituye una noción *explicativa* (es decir, una noción *explicativa causal*). Como nuestra intención es que la existencia de las diversas valencias figure en la química como *causas*, debemos señalar qué *es* la valencia, y no limitarnos a dar sus valores numéricos. Pero Leeds *niega* que la *referencia* sea una noción explicativa causal. Necesitamos nociones como verdad y referencia para *expresar* ciertas cosas (que podrían, en principio, ser explicadas por otros medios, como el uso de lenguajes infinitarios). Para *este* propósito, es *insustancial* si la referencia primitiva se define de la manera que Field llama 'absurda'. La referencia no es (o, en todo caso, Field no ha demostrado que sea) una noción explicativa causal. En suma, podemos dar un 'argumento trascendental' para el procedimiento de Tarski si recurrimos a un *propósito* para la posesión de nociones como *verdad* y *referencia* que no presente ningún parecido con el propósito por el cual tenemos nociones como *valencia*.

Se podría objetar que Leeds va a encontrar dificultades con la verdad en el *meta*-lenguaje. Pero se aplica el mismo procedimiento, salvo en que el metalenguaje debe tener funciones de verdad infinitas —$<_2$ — *más largas*. De cualquier modo, el punto crucial sobre los lenguajes infinitarios no es realmente esencial para el argumento de Leeds. Lo importante consiste en justificar el criterio *T* señalando una razón, distinta de la explicación causal, para tener nociones como *verdad* y *referencia* en el lenguaje, y mostrando que satisfacer el criterio *T* nos da la clave.

Un problema más profundo del argumento de Leeds es éste: ¿qué hacemos con enunciados como

(4) Lo que él dijo era verdadero

cuando el sujeto al que aludimos no hablaba nuestro lenguaje objeto? Una línea natural de razonamiento es que la noción de referencia *sólo* tiene sentido cuando la noción de traducción lo tiene. Si no podemos traducir *en absoluto* lo que 'él' estaba diciendo (desde luego, en la mayoría de los casos no se requiere que la traducción capte cada matiz para mostrarnos los valores de verdad de lo que fue dicho), ¿cómo podríamos *comprender* una expresión como (4)?

Lo anterior sugiere la siguiente idea: pensemos *primero* en la referencia como si estuviera definida para nuestro lenguaje al estilo de Tarski, y extendámosla entonces a otros lenguajes *por medio de* una traducción. Field objetaría, por otro lado, que si me estoy refiriendo a X ello significa (o consiste en, o quizás es reductible empíricamente a —en el sentido de reducción empírica que posee 'el agua es H_2O') que me encuentro en cierta relación *fisicalista* con X, tentativamente la relación de hallarse conectado a X o con alguna propiedad de X, a través de cierta clase de cadena causal (¡que se debe especificar!). Y si 'él' se refiere a X, al margen del lenguaje que hable, se encuentra en la misma relación fisicalista con X. Así, lo que *significa* para él *referirse* a X *debe* ser explicado sin referirse a *mi* lenguaje, y sin emplear el concepto de *traducción*. Pero, ¿por qué? Es obvio que podemos justificar la elaboración de un mapa de los lenguajes de otros pueblos a partir del nuestro sin presuponer que existe una relación de referencia unitaria (ya no digamos fisicalista) entre las palabras y las cosas: hacer mapas de lenguajes extraños a partir de lenguajes familiares (cuando es posible) constituye una forma de *racionalizar* la conducta de otras personas, y por consiguiente de facilitar el trato con ellas, predecir su conducta, etcétera.

Y si se nos justifica (por el razonamiento de Leeds) el poseer una noción de referencia para nuestro propio lenguaje (una noción 'localista' de referencia, como diría Quine), estamos justificados para extenderla a cualquier otro lenguaje del que elaboremos un mapa partiendo del nuestro, de acuerdo con el método que usemos en su elaboración.

Ahora bien, nada de lo expuesto demuestra que Field se halle en un error. *Quizás* exista una relación fisicalista, de la clase que Field considera, entre una palabra en L (donde 'L' es una *variable* sobre los lenguajes naturales) y cualquier x a la que se refiera, sin importar el hecho de que no necesitamos presuponer o creer esto

para justificar la introducción de expresiones como 'verdadero' y 'se refiere' a nuestro lenguaje. Pero la obligación de probarlo parece corresponder en gran medida a Field.

Por otro lado, creo que no es *totalmente* incorrecto pensar la referencia como una noción explicativa causal. En la siguiente conferencia empezaré a desentrañar una explicación que le hace justicia a esta intuición realista.

SEGUNDA CONFERENCIA*

Aunque sin duda resulta beneficioso que las palabras terminadas en 'ismo' hayan pasado de moda en filosofía, *algunas* parecen presentar una notable resistencia a desaparecer. Una de ellas es 'realismo'. Cada vez son más los filósofos que hablan en la actualidad sobre realismo, pero se dice muy poco acerca de qué es éste.

Al margen de otras cosas que postulan, los realistas sostienen por lo regular que creen en una 'teoría de la verdad por correspondencia'.

Cuando los realistas *defienden* su posición, normalmente lo hacen *impugnando* alguna versión del idealismo, que en nuestro tiempo, podría ser el positivismo o el operacionalismo. (Esto no es sorprendente, pues todos los filósofos intentan transferir a sus oponentes la obligación de la prueba. Y si éstos poseen tal obligación, refutar sus afirmaciones parece suficiente defensa de la propia posición.) Y el argumento característico de los realistas contra el idealismo es que convierte en un *milagro* el éxito de la ciencia. Berkeley necesitaba a Dios sólo para dar razón de la aceptación de las creencias sobre mesas y sillas (y los árboles de la explanada); pero la apelación a Dios ha pasado de moda en filosofía, y en cualquier caso el uso que le dio Berkeley es muy extraño desde el punto de vista de la mayor parte de los deístas. Y los positivistas modernos se ven forzados a dejar sin explicación (según los acusa el realista) el hecho de que los 'cálculos de electrones', los 'cálculos del espacio-tiempo' y los 'cálculos del ADN'

* La Segunda conferencia y la primera sección de la Tercera conferencia son una versión ampliada de "What is 'Realism'?" (© 1976, The Aristotelian Society), documento presentado ante la Sociedad Aristotélica el 23 de febrero de 1976. Agradezco al editor de la Sociedad Aristotélica por haberme permitido la reimpresión del texto.

predigan correctamente fenómenos observables a pesar de que, en realidad, no existen los electrones, ni el espacio-tiempo curvo, ni las moléculas de ADN. Si tales cosas existieran, una manera natural de dar cuenta del éxito de estas teorías consistiría en que se trata de *explicaciones parcialmente verdaderas* de cómo se comportan. Y una razón natural de la forma en que las teorías científicas se suceden entre sí —digamos, cómo la relatividad de Einstein llegó a ser sucesora de la gravitación universal de Newton—, es que una explicación parcialmente correcta o incorrecta de un objeto teórico —como el campo gravitacional o la estructura métrica del espacio-tiempo o ambos—, se sustituye por una explicación *mejor* del mismo objeto u objetos. Pero si éstos no poseen existencia real, entonces es un *milagro* que una teoría que habla de la acción gravitacional a distancia prediga fenómenos con éxito; es un *milagro* que una teoría que habla de un espacio-tiempo curvo prediga fenómenos con éxito; y el hecho de que las leyes de la teoría anterior sean derivables 'en el límite' de las leyes de la posterior no tiene ninguna significación metodológica.

No estoy afirmando que el positivista (o lo que sea) carezca de *réplicas* para este tipo de argumentos. Posee varias: las teorías reduccionistas sobre el *significado* de los términos teóricos, las teorías de la explicación, etc. Mi interés aquí recae más bien en el siguiente hecho: el argumento del realista depende del éxito de la ciencia o, en tiempos pasados, del éxito de una teoría del sentido común sobre los objetos materiales. Pero ¿cómo se relaciona el éxito de la ciencia con la teoría de la verdad por correspondencia o, para el caso, con *cualquier* teoría de la verdad?

Es un hecho empírico indudable que la ciencia logra hacer muchas predicciones verdaderas, proyectar mejores formas de controlar a la naturaleza, etc. Si el realismo constituye una *explicación* de este hecho, debe ser él mismo una *hipótesis* científica totalizadora. Sus adherentes han abrazado a menudo esta idea proclamando que el realismo *es* una hipótesis empírica.[1] Pero entonces no queda claro

[1] Resultará claro que soy de la opinión de que el realismo *se parece* a una hipótesis empírica en que (1) puede ser falso y (2) los hechos son importantes para su defensa (o su crítica); pero ello no implica que sea *científico* (en ninguna de las acepciones comunes del término) ni que constituya una *hipótesis*. La noción de 'científico' será discutida más adelante en estas conferencias. En artículos previos he expuesto por qué creer en un mundo externo y en otras mentes no es una 'hipótesis' (véanse los capítulos 1 y 17 de mi libro *Mind, Language, and Reality; Philosophical Papers*, vol. 2, Cambridge University Press, 1975).

cómo se halla relacionado con la teoría de la *verdad*. En esta conferencia trataré de mostrar cuál es la conexión entre explicar el éxito del conocimiento y la teoría de la verdad.

1. *La 'convergencia' del conocimiento científico*

Con frecuencia sus proponentes denominan 'realismo científico' a lo que llamo 'realismo'. Si evito tal término aquí se debe a que la etiqueta de 'realista científico' conlleva un cierto *matiz* ideológico, un matiz bastante reminiscente del materialismo decimonónico o, para decirlo sin ambages, del ateísmo provinciano. De hecho, si 'realista científico' es quien cree, *inter allia*, que *todo* conocimiento digno de ese nombre forma parte de la 'ciencia', entonces yo no lo soy. Pero sin duda el conocimiento científico representa una parte sustancial de nuestro saber, y su naturaleza y significación han interesado a todos los grandes filósofos preocupados de algún modo por la epistemología. Así que no es de sorprender que tanto realistas como idealistas se proclamen 'filósofos de la ciencia', en *dos* sentidos de la preposición 'de'. Y si en lo que sigue me centro en el conocimiento científico, se debe a que el debate gira a su alrededor y no por un compromiso personal con el cientificismo.

Para comenzar, permítaseme señalar que, según creo, *hay* algo de cierto en la idea de *convergencia* del conocimiento científico. *En qué consiste* esto se explica de un modo excelente, según mi opinión, en un artículo todavía inédito de Richard Boyd.[2] Éste observa que todo lo que puede concluirse de la filosofía científica convencional (positivista) es que las últimas teorías de una ciencia, para que resulten *mejores* que las precedentes, deben hacer alusión a muchas de las *oraciones observacionales* de éstas (especialmente las oraciones observacionales *verdaderas* implicadas por las teorías previas). De esto no se sigue que las teorías más recientes deban implicar la *verdad aproximada de las leyes teóricas de las teorías anteriores bajo ciertas circunstancias* —algo que se hace a menudo. De hecho, preservar en lo posible los *mecanismos* de la teoría previa, que es lo que los científicos intentan hacer (o mostrar que se trata de 'casos límite' de nuevos mecanismos), constituye con frecuencia la manera *más difícil* de construir una teoría capaz de mantener las viejas predicciones observacionales, cuando son correctas, y de incorporar si-

[2] *Realism and Scientific Epistemology*, Cambridge University Press, en prensa.

multáneamente nuevos datos de la observación. Es un hecho que los científicos tratan de proceder así, por ejemplo, *preservan* si pueden la conservación de la energía antes que postular violaciones a ella, y también es un hecho que esta estrategia ha conducido a importantes descubrimientos (desde Neptuno hasta el positrón).

Boyd busca describir al realismo como una hipótesis empírica totalizadora mediante dos principios:

(1) Los términos de una ciencia desarrollada normalmente *refieren*.

(2) Las leyes de una teoría perteneciente a una ciencia desarrollada son por lo general aproximadamente *verdaderas*.

El propósito de su ensayo es mostrar que los científicos actúan de este modo porque *creen* en (1) y en (2) y que su estrategia funciona porque (1) y (2) son *verdaderos*.

Una de las cosas más interesantes de este argumento es que, de ser correcto, las nociones de 'verdad' y 'referencia' desempeñan una función explicativa causal en la epistemología. (1) y (2) son premisas dentro de una *explicación* del comportamiento de los científicos y del éxito de la ciencia, y contienen en esencia conceptos tomados de la semántica referencial. Si se remplaza 'verdaderas' en la premisa (2) (desde luego, el argumento de Boyd necesita de muchas más premisas que (1) y (2)), por algún 'sustituto' operacional —por ejemplo, 'son simples y conducen a predicciones verdaderas'— no se preserva la explicación.

Detengámonos a ver por qué. Supongamos que T_1 es la teoría aceptada en alguna rama central de la física (la *física* se cuenta seguramente entre las ciencias 'desarrolladas', si existe alguna), y que yo soy un científico que intenta encontrar una teoría T_2 que remplace a T_1. (Quizás incluso conozco áreas en las que T_1 puede llevar a predicciones falsas.) Si creo en los principios (1) y (2) entonces sé que las leyes de T_1 son (probablemente) verdaderas de un modo aproximado. Así que T_2 debe poseer una cierta propiedad —la propiedad de que las leyes de T_1 sean 'verdaderas por aproximación' *juzgadas desde el punto de vista de* T_2— o T_2 no tendrá (probablemente) posibilidades de ser verdadera. Como quiero teorías que no *sólo* sean 'verdaderas por aproximación', sino teorías que tengan la posibilidad de ser *verdaderas*, me restringiré a considerar, como candidatas para T_2, a teorías que presenten esta propiedad, es decir, que contengan

las leyes de T_1 como un caso límite. Pero tal es, justamente, la característica del método científico que hemos expuesto. (Boyd también explica muchas otras características del método científico, y no sólo este aspecto de la 'convergencia', pero no necesito abordarlas aquí.) En suma, mi conocimiento de la verdad de (1) y (2) me permite circunscribir la clase de teorías posibles que debo considerar, y por lo tanto incrementa mis perspectivas de éxito.

Ahora bien, si todo lo que sé es que T_1 lleva (principalmente) a predicciones verdaderas en algún vocabulario observacional (una noción que he criticado en otro lugar),[3] entonces lo único que sé de T_2 es que debería abarcar a la mayoría de las 'oraciones observacionales' implicadas por T_1. Pero de ahí *no* se concluye que también deba incluir la verdad de las *leyes* de T_1 en determinado límite. Hay muchas otras maneras de construir T_2 de modo que ésta abarque a la verdad de la mayoría de las oraciones observacionales de T_1 ; y hacer que T_2 implique la 'verdad aproximada' de las *leyes* de T_1 es a menudo el camino *más difícil*. Tampoco hay ninguna razón por la que T_2 deba tener la propiedad de que podamos asignarle *referentes* a los términos de T_1 desde el punto de vista de T_2. Sin embargo, es un hecho que podemos asignar un referente al 'campo gravitacional' de la teoría newtoniana *desde el punto de vista* de la teoría de la relatividad (aunque no al 'éter' ni al 'flogisto'); un referente al 'gene' de Mendel desde el punto de vista de la biología molecular de hoy, y un referente al 'átomo' de Dalton desde el punto de vista de la mecánica cuántica. Estas asignaciones retrospectivas de referencia dependen de lo que ha sido llamado el 'principio del beneficio de la duda' o el 'principio de caridad',[4] pero no de una 'caridad' *irrazonable*. Indudablemente el 'gene' del que habla la biología molecular es el gene (o más bien el 'factor') al que Mendel *quería* aludir; ciertamente se trata de lo que debió querer decir. Repito que en caso de creer que los términos de T_1 tienen referentes (y en mi supuesto, mi teoría semántica incorpora el principio del beneficio de la duda), entonces el hecho de que T_2 deba poseer esta propiedad, la de que *desde su punto de vista* sea capaz de asignar referentes a los términos de T_1, representará una *restricción*, es decir, limitará la clase de teorías que

[3] En 'What Theories are Not', reimpreso en mi *Mathematics, Matter, and Method; Philosophical Papers*, vol. 1, Cambridge University Press, 1975.

[4] En 'Language and Reality' en mi libro *Mind, Language, and Reality; Philosophical Papers*, vol. 2, Cambridge University Press, 1975.

pueden incorporarse a ella. Y, nuevamente, si no aplico las nociones de verdad y referencia a la filosofía de la ciencia, si únicamente utilizo propiedades 'globales' del orden de 'simplicidad' y 'conduce a predicciones verdaderas', careceré de una analogía para esta restricción y no me será posible delimitar la clase de teorías que podrían ser T_2 de esta forma.

2. *¿Qué pasaría si no hubiera 'convergencia' en el conocimiento científico?*

Permítaseme abordar ahora estos temas desde el otro extremo, desde el problema de la 'verdad'. ¿Cómo se verían afectadas nuestras nociones de *verdad* y *referencia* si decidiéramos que *no hay* convergencia en el conocimiento científico?

Ésta es ya la situación según algunos, como Kuhn, que se muestra escéptico con respecto a la convergencia y escribe (por lo menos en *La estructura de las revoluciones científicas*) como si el mismo término no pudiera tener el mismo referente en paradigmas distintos (afirma que teorías que pertenecen o generan paradigmas diferentes corresponden a 'mundos' diversos), y esto es aún más pronunciado en la opinión de Feyerabend.

Supongamos que tienen razón y que el 'electrón' de la teoría de Bohr (la teoría de Bohr y Rutherford formulada a principios del siglo XX) no se refiere a lo que *ahora* nosotros llamamos electrones. Entonces no se refiere a *nada* que podamos reconocer en la teoría actual y, más aún, no se refiere a *nada desde el punto de vista* de la teoría actual (desde este punto de vista, Bohr sólo pudo haberse referido a los electrones, y si no fue así, no se refirió a nada). Por consiguiente, si recurrimos a la teoría actual para contestar a la pregunta "¿hacía Bohr una referencia al emplear el término 'electrón'?", la respuesta tendría que ser 'no', según Kuhn y Feyerabend. ¿Y qué otra teoría podríamos utilizar sino nuestra teoría presente? (Se podría llamar a esto el predicamento kantiano, aunque Quine también es muy aficionado a él.) Kuhn habla como si cada teoría tuviera referencias —a saber, a su *propio* 'mundo' de entidades—, pero ello no es así para *ninguna* teoría (científica).

Feyerabend llega a su posición por el siguiente razonamiento (con el que Kuhn no está de acuerdo en absoluto; cualquier semejanza en sus concepciones sobre la referencia transteórica no deriva de un análisis compartido de la ciencia): quien introduce un término

científico, o los especialistas que lo usan, aceptan ciertas leyes como verdades virtualmente necesarias sobre el referente putativo. En efecto, Feyerabend trata a esas leyes, o a la descripción teórica del referente basada en esas leyes, como una *definición* del referente (de hecho como una definición *analítica*). En consecuencia, si alguna vez decidimos que no existe nada que se ajuste a esa *descripción* exacta, deberemos decir que 'no había tal cosa'. Si nada se ajustaba a la descripción exacta del electrón de Bohr y Rutherford, el 'electrón' *en el sentido que lo usaron Bohr y Rutherford* no posee referente. Aún más: si su descripción teórica varía en dos teorías, el término 'electrón' tiene un *sentido* diferente en cada una de ellas (por cuanto se trata de un sinónimo con descripciones distintas; Feyerabend no lo expresa de manera explícita, pero resulta obvio que éste es su argumento). En general, concluye Feyerabend, un término no puede tener un referente compartido ni un sentido compartido en teorías diferentes (la 'inconmensurabilidad de las teorías').

Esta línea de razonamiento puede ser refutada (como lo hemos hecho Saul Kripke y yo en varios lugares) mediante el argumento de que término científico y descripción no son sinónimos. Inclusive, uno de los principios esenciales de la metodología semántica es que cuando los hablantes especifican con una *descripción* al referente de un término que utilizan, y aquélla no cumple con la referencia debido a las creencias fácticas errónes de los hablantes, debemos presuponer que aceptarían alguna reformulación razonable de su descripción (en los casos en que, dado nuestro conocimiento, tuviéramos en claro cómo reformular su descripción para que satisfaga la referencia, y no hubiera ninguna ambigüedad con respecto a la forma de efectuarlo en un contexto práctico). (Éste es, a grandes rasgos, el principio del beneficio de la duda ya mencionado.)

Para dar un ejemplo: no hay nada en el mundo que encaje *con exactitud* en la descripción de Bohr y Rutherford de un electrón. Pero existen partículas que se adecuan *de manera aproximada* a la descripción de Bohr: presentan la carga y la masa correctas y son responsables de efectos determinantes que Bohr y Rutherford explicaron en términos de 'electrones'; la corriente eléctrica en un alambre constituye, por ejemplo, un flujo de esas partículas. El principio del beneficio de la duda nos obliga a considerar que Bohr se refería a ellas.

Por cierto, si Bohr no hubiera concedido el beneficio de la duda a su posición anterior (el periodo Bohr-Rutherford), no habría *seguido*

utilizando el término 'electrón' (¡sin un solo comentario!) cuando en los años treinta participó en la creación de la mecánica cuántica.

Pero volvamos a Kuhn: podemos responderle que *existen* entidades —de hecho, las que justamente hoy llamamos 'electrones'— que se comportan en muchos de sus aspectos como los 'electrones' de Bohr (hay uno por cada átomo de hidrógeno, tienen carga negativa, la masa apropiada, etc.). Y el principio del beneficio de la duda (por supuesto, esto equivale solamente a responder a Kuhn de la misma manera en que lo hicimos con Feyerabend) dicta que en estas circunstancias debemos entender que Bohr se refirió a lo que llamamos 'electrones'. Sólo necesitamos agregar que poseemos una teoría distinta sobre las *mismas* entidades a las que Bohr llamó 'electrones' en ese entonces; su término tenía en efecto un referente.

Pero podemos razonar así sólo porque la teoría actual afirma la existencia de entidades que cumplen muchas de las *funciones* que supuestamente desempeñaban los 'electrones' de Bohr, aun cuando estas entidades presentan otras propiedades muy peculiares, como la complementaridad de posición y cantidad de movimiento, que no debían observarse en los 'electrones' de Bohr y Rutherford. Pero ¿qué sucedería si aceptáramos una teoría que postulase que los electrones son como el *flogisto*?

Tendríamos que decir entonces que los electrones realmente no existen. ¿Y si esto continuara ocurriendo? ¿Y si *todas* las entidades teóricas formuladas por una generación (moléculas, genes, etc., además de electrones), 'no existieran' desde el punto de vista de la ciencia posterior? Desde luego, se trata de otra forma del viejo y escéptico 'argumento del error': ¿cómo saber si uno no se encuentra *ahora* en un error? Pero, bajo esta forma, el argumento del error constituye hoy una *seria* preocupación para mucha gente y no sólo una 'duda filosófica'.

Una causa de que represente una seria preocupación consiste en que eventualmente la siguiente metainducción se vuelve obligada de un modo abrumador: *puesto que ningún término utilizado en la ciencia durante los últimos cincuenta años* (o los que sean) *tenía referente, hay que concluir que tampoco lo tiene ninguno de los términos utilizados ahora* (excepto tal vez los términos observacionales, si es que existen).

Una aspiración obvia de la teoría de la referencia es refutar la metainducción citada, lo cual constituye una de las justificaciones del principio del beneficio de la duda. Pero éste quizás *no resulte*

razonable; nosotros no lo mantendremos hasta el grado de llegar a decir que el *flogisto* tenía un referente. Si no existe la convergencia, si las teorías científicas posteriores ya no toman a las previas como 'casos límite', si los principios (1) y (2) de Boyd son claramente falsos desde el punto de vista de la ciencia futura, entonces el beneficio de la duda carecerá siempre de sentido: no se podrá *modificar* de manera razonable a las descripciones que las teorías previas hicieron de diversas entidades, para referirlas a las otras entidades, con funciones más o menos iguales, que existen desde el punto de vista de las teorías más recientes. La referencia se derrumbaría.

Pero ¿qué pasa con la noción de *verdad* en la ciencia teórica si ninguno de los términos descriptivos tiene referentes? Tal vez todas las oraciones teóricas sean 'falsas', o se asuma alguna convención para asignar valores de verdad cuando los predicados no poseen referentes. En cualquier caso, la noción de 'valor de verdad' carecerá de interés para las oraciones que contengan términos teóricos. Así también la verdad se derrumbaría.

Quisiera argumentar ahora que lo precedente *no es* en realidad lo que pasaría. Aunque ello nos conducirá a consideraciones lógicas muy sutiles.

3. *Intuicionismo matemático: una aplicación al conocimiento empírico*

Bajo el supuesto de que el lector no estudió el 'intuicionismo matemático' (la escuela de filosofía matemática desarrollada por Brouwer, Heyting, etc.), me tomaré la libertad de mencionar algunos datos que voy a utilizar en lo que sigue.

Una idea intuicionista esencial consiste en el empleo de las conectivas lógicas en un sentido que se aparta de lo clásico. (Proceden así, desde luego, porque consideran que la acepción 'clásica' es inaplicable al razonamiento de dominios infinitos o potencialmente infinitos.) Explican este sentido —es decir, explican *sus* significados para las conectivas lógicas— en términos de *demostrabilidad* constructiva más que en los términos (clásicos) de verdad.

Así:

(1) Afirmar *p* es afirmar *p es demostrable*. ('*p* · ⌈*p* no es demostrable⌉' es para los intuicionistas una *contradicción*.)

(2) '$\neg p$' ($\neg$ es el símbolo de los intuicionistas para la negación) significa: *es demostrable que una prueba de p implique la demostrabilidad de* $1 = 0$ (o cualquier otro absurdo evidente). En otras palabras, $\neg p$ afirma *el absurdo de la demostrabilidad de p* (y no la 'falsedad' clásica de p).

(3) '$p \cdot q$' significa: *p es demostrable y q es demostrable.*

(4) '$p \vee q$' significa: *existe una prueba de p o una prueba de q y se puede sabel cuál es.*

(5) '$p \supset q$' significa: *existe un método que una vez aplicado a cualquier prueba de p lleva a una prueba de q (y una prueba de el método realiza esto).*

Estos significados son claramente distintos de los clásicos. Por ejemplo, $p \vee \neg p$ (que afirma la determinabilidad de toda proposición) no es un teorema del cálculo proposicional intuicionista.

Ahora *reinterpretemos* las conectivas *clásicas* de la siguiente manera:

(1) $\sim$ es idéntico a $\neg$.

(2) $\cdot$ (clásico) se identifica con $\cdot$ (intuicionista).

(3) $p \vee q$ (clásico) se identifica con $\neg(\neg p \cdot \neg q)$.

(4) $p \supset q$ (clásico) se identifica con $\neg(p \cdot \neg q)$.

Pero entonces, con esta interpretación, los teoremas del cálculo proposicional *clásico* se convierten en teoremas del cálculo proposicional intuicionista.[5] En otras palabras, ésta es una *traducción* del cálculo proposicional clásico *al* cálculo proposicional intuicionista, aunque por supuesto, no en el sentido de presentar los *significados* clásicos de las conectivas según las nociones intuicionistas, sino en el de presentar los teoremas clásicos. (Por cierto, no es la única 'traducción' de este tipo.) Si las conectivas se reinterpretan así, los significados siguen hallándose fuera de lo clásico, pues se explican en términos de *demostrabilidad* y no de *verdad y falsedad.*

Para ejemplificar: desde el punto de vista clásico, $p \vee \sim p$ afirma que toda proposición es verdadera o falsa. Con la 'traducción de la negación y la conjunción' a la lógica intuicionista ya citada, $p \vee \sim p$

[5] Esto fue señalado por Kurt Gödel en 'On Intuitionistic Arithmetic and Number Theory', reimpreso en Martin Davis (ed.), *The Undecidable*, Nueva York, Raven Press, 1965.

afirma $\neg(\neg p \cdot \neg\neg p)$, lo cual implica que resulta absurdo que una proposición y su negación sean ambas absurdas, y no dice nada acerca de su verdad o falsedad.

Se puede extender lo anterior a los cuantificadores, pero prefiero omitir los detalles.

Una de las cosas que esto pone de manifiesto, y que se opone a lo que han afirmado muchos filósofos —entre quienes, desde hace poco, se incluye Hacking—, es que las reglas de inferencia, como $p \cdot q / \therefore p$; $p \cdot q / \therefore q$; $p / \therefore p \vee q$; $q / \therefore p \vee q$; $\sim p, \sim q / \therefore \sim (p \vee q)$, *no fijan los 'significados' de las conectivas lógicas. Se podrían aceptar todas estas reglas (y también todas las tautologías clásicas) y seguir usando las conectivas lógicas en el sentido contrario al clásico que recién describimos, un sentido que no es veritativo-funcional.*

Supongamos ahora que aplicamos *esta* interpretación de las conectivas lógicas (la interpretación mencionada arriba que se efectúa mediante la 'traducción de la conjunción y la negación') a la ciencia empírica, de la siguiente manera (esta idea se me ocurrió mientras leía la conferencia de Dummett sobre la verdad, pero no debe atribuírsele la responsabilidad de la misma): remplacemos la *demostrabilidad constructiva* (en el sentido del intuicionismo matemático) por la *demostrabilidad constructiva derivada de* (una reconstrucción coherente y adecuada de) *los postulados pertenecientes a la ciencia empírica aceptada en ese momento*[6] (o, si se quiere ser realista con respecto a los 'enunciados observacionales', de aquellos que acompañan al conjunto de enunciados observacionales verdaderos). Si esta ciencia es en sí misma *incongruente* con el conjunto de enunciados observacionales verdaderos —porque implica una predicción falsa—, habría que especificar un subconjunto apropiado, pero no consideraré aquí cómo hacerlo. Si B_1 es la ciencia empírica aceptada en *una* época y B_2 es la ciencia empírica aceptada en una época

[6] Uno de los problemas para reformular la física en términos intuicionistas consiste en que no se pueden afirmar leyes como la de la gravitación de Newton, según la cual dos números reales diferentes, determinados empíricamente, (la fuerza de A sobre B y el cociente de g veces el producto de las masas y la distancia al cuadrado) son exactamente iguales en la matemática intuicionista (para una explicación de esto, *cfr.* mi artículo 'What is Mathematical Truth?', en *Mathematics, Matter, and Method; Philosophical Papers*, vol. 1, Cambridge University Press, 1975). Pero *puede* decirse que tal ley se mantiene para, por ejemplo, treinta lugares decimales, y si de cualquier modo no se espera que la ley se mantenga a largo plazo ni se pretende 'converger', mediante aproximaciones sucesivas, a algo que sea objetivamente *verdadero*, entonces cabe suponer que no importa que la teoría física *se debilite* hasta ese extremo.

diferente, entonces, de acuerdo con esta interpretación 'cuasi-intuicionista', las mismas *conectivas lógicas* se referirían a la 'demostrabilidad en B_1 ' cuando se utilizaran en B_1 y a la 'demostrabilidad en B_2 ' cuando se utilizaran en B_2 . Las *conectivas lógicas* modificarían su significado de manera sistemática al cambiar el conocimiento empírico.

4. *Verdad*

Supongamos que formalizamos la ciencia empírica o alguna parte de la ciencia empírica, es decir, que la formulamos en un lenguaje formalizado L, con reglas lógicas y axiomas convenientes, y con postulados empíricos apropiados al cuerpo de la teoría que estamos formalizando. Siguiendo las prácticas lógicas habituales del presente, el predicado 'verdadero' (aplicado a oraciones de L) podría no ser en sí mismo un predicado de L, sino pertenecer a un metalenguaje más enfático, ML. (Saul Kripke investiga actualmente un método para evitar esta separación entre lenguaje objeto y metalenguaje, pero el mismo no afectaría a la presente exposición.) Este predicado podría ser definido (utilizando los recursos lógicos de ML pero sin otro vocabulario descriptivo que el de L) por los métodos que debemos a Tarski; o podría tomarse como una noción primitiva (no definida) de ML. En cualquier caso, esperaríamos que fueran teoremas de ML todas las oraciones de la conocida forma:

(T) 'La nieve es blanca' es verdadera si y sólo si la nieve es blanca

a saber, todas las que afirmen la equivalencia de una oración de L (aseguramos que 'La nieve es blanca' es un enunciado de L) y la oración de ML que dice que ese enunciado es verdadero. (En su *Wahrheitsbegriff* Tarski llamó a esto 'criterio W', y por alguna razón fue traducido como *'convención T'*. Por mi parte, denominaré *criterio T* al requisito de que todas las oraciones de la forma (T) sean teoremas de ML.)

¿Qué sucede con el término 'verdadero' si reinterpretamos las conectivas lógicas en la forma 'cuasi intuicionista' que acabamos de describir? *Es posible definirla exactamente al estilo de Tarski.* Sólo que 'verdad' se convierte en *demostrabilidad* (o, para ser más preciso, en la doble negación de la demostrabilidad. Ignoraré esta última sutileza). En resumen: la propiedad *formal* de la verdad —el criterio de

adecuación (criterio T)— se limita a *fijar* la extensión del término 'verdadero' *si las conectivas lógicas son las clásicas*.

Ello significa que podemos ampliar la observación que hicimos en la sección 3 (el primer párrafo en cursivas): aun si los 'nativos' que estamos estudiando aceptaran el criterio T además de aceptar todas las tautologías clásicas, de aquí no se desprende que su 'verdadero' sea el 'verdadero' clásico.

La 'verdad' (definida, siguiendo a Tarski, de la manera recursiva habitual) se convierte en *demostrabilidad* si las conectivas lógicas se reinterpretan apropiadamente. ¿Pero en qué se convierte entonces la 'referencia'?

En la definición de verdad y referencia de Tarski,

(a) 'Electrón' tiene referencia

equivale a

(b) Existen electrones.

Pero si 'existen' se interpreta de manera intuicionista, (b) sólo afirma que

(c) Existe una descripción D tal que 'D es un electrón' es demostrable en B_1.

Y *esto* puede ser verdadero (para un B_1 adecuado) aunque no existieran electrones. En suma, el efecto de reinterpretar de manera intuicionista las conectivas lógicas es que la 'existencia' se vuelve *intrateórica*. De hecho, el efecto resulta más complicado que (c) si, además de entender las conectivas de un modo 'cuasi intuicionista' (es decir, en la forma intuicionista, pero con la 'demostrabilidad' relativizada a B_1), utilizamos la traducción de la conjunción y la negación para interpretar las conectivas 'clásicas', como aquí se ha sugerido. Pero tal complejidad no cambia el punto que acabamos de mencionar: si los cuantificadores, como las otras conectivas lógicas, se interpretan de acuerdo con la noción de *demostrabilidad*, entonces se vuelven intrateóricos.

5. *Teoría de la verdad por correspondencia*

Me gustaría proponer ahora (probablemente el lector se pregunte adónde quiero llegar) que el efecto de abandonar el realismo —es decir, dejar de lado la creencia de que existe un mundo descriptible

de cosas inobservables, y aceptar en cambio la creencia de que todas las 'cosas inobservables' (y, posiblemente, también las observables) de las que se habla en una generación de teorías científicas, incluyendo la nuestra, son *meras* conveniencias teóricas, destinadas a ser remplazadas y suplantadas en el futuro por construcciones teóricas distintas—, *no* significaría descartar totalmente los predicados *verdadero* y *se refiere* en sus aspectos *formales*. Como lo indica la exposición anterior, podríamos *conservar* la semántica formal (e incluso las definiciones de verdad 'tipo Tarski'), conservar también la lógica clásica, y no obstante *transmutar* nuestra noción de 'verdad' en algo que se aproxime a la 'aseverabilidad justificada'. Y creo que en efecto esta transmutación llegaría a producirse. (Desde luego, los detalles formales sólo son una reconstrucción racional, y no la única posible.)

Por supuesto, no hay manera de *probar* tal afirmación. Se trata de una especulación acerca de la naturaleza cognoscitiva humana, la cual se expresa bajo la forma de una predicción sobre una situación hipotética. Pero lo que la vuelve plausible es que justamente tal sustitución —el remplazo de la noción realista de verdad por la 'verdad dentro de la teoría' o la 'aseverabilidad justificada'— *siempre* ha acompañado al escepticismo concerniente a la noción realista, desde Protágoras hasta Michael Dummett.

Si esto es correcto, ¿cuál es la respuesta a nuestra pregunta original acerca de qué relación existe entre las explicaciones realistas del método científico, su éxito, su convergencia, y la concepción realista de la verdad?

Al principio observamos que los realistas aseguran creer en algo llamado teoría de la verdad por correspondencia. ¿Pero en qué consiste ésta?

Si estoy en lo cierto, no se trata de una *definición* distinta de la verdad. Sólo existe un método que todos conocen para *definir* lo 'verdadero', y es el de Tarski. (En realidad, como ya mencionamos, Saul Kripke encontró un método *nuevo*, pero su diferencia con el de Tarski no es esencial en este contexto, aunque sí para tratar las antinomias.) Pero el método de Tarski, ¿es 'realista'?

Bueno, ello depende. Si las conectivas lógicas son entendidas de forma realista ('clásica', al decir de la gente), una definición de verdad al estilo de Tarski es 'realista' al menos en este sentido: la satisfacción (de la cual la verdad representa un caso especial) constituye una relación entre palabras y cosas o, más precisamente, en-

tre fórmulas y secuencias finitas de cosas. ('Satisface' es el término técnico que Tarski utiliza para lo que he estado llamando *referencia*. Por ejemplo, en lugar de decir " 'Electrón' se refiere a los electrones", él afirmaría: "La secuencia de longitud uno que consiste únicamente de x satisface la fórmula 'Electrón (y)' si y sólo si x es un electrón". 'Satisface' posee la ventaja técnica de aplicarse a fórmulas de situación n. Por ejemplo, se puede decir que la secuencia *Abraham; Isaac* satisface la fórmula 'x es el padre de y'; pero no se acostumbra utilizar 'se refiere' con relación a fórmulas diádicas, etc., como decir que 'padre de' se refiere a *Abraham; Isaac*.) Ciertamente ello concuerda con una parte esencial de la idea de una teoría de la correspondencia.

Sin embargo, uno tiende a sentirse inconforme con la teoría de Tarski en tanto reconstrucción de la 'teoría de la verdad por correspondencia' *aun cuando* las conectivas lógicas se entiendan bajo la forma clásica. Pienso que existe un buen número de razones para esta inconformidad, que he expresado en algunos de mis escritos, aunque creo que Hartry Field encontró la principal: el hecho de que la referencia primitiva (es decir, la *satisfacción* en el caso de los predicados primitivos del lenguaje) se 'explique' mediante una *lista* contituye la mayor causa de contrariedad.

Pero la lista tiene una *estructura* muy especial. Observemos las cláusulas siguientes desde el punto de vista de la definición de la referencia primitiva:

(1) 'Electrón' se refiere a los electrones.

(2) 'Gene' se refiere a los genes.

(3) 'Molécula de ADN' se refiere a las moléculas de ADN.

Las mismas se parecen a la famosa

(4) 'La nieve es blanca' es verdadera si y sólo si la nieve es blanca

y la similitud no es casual: 'verdadera' es el caso O-ádico de satisfacción (una fórmula es verdadera si no tiene variables libres y la secuencia cero la satisface). El criterio de adecuación (el criterio T) puede generalizarse de esta forma:

> (Llamemos al resultado 'criterio S' —'S' por satisfacción): una definición adecuada de satisfacción para L debe reconocer como teoremas a todas las instancias del siguiente esquema:

'$\lceil P(x_1, \ldots, x_n) \rceil$' se satisface con la secuencia $y_1, \ldots, y_n$ si y sólo si $P(y_1, \ldots, y_n)$.

Si reescribimos (1) como

(1') 'Electrón (x)' es satisfecho por y_1 si y sólo si y_1 es un electrón

que es la forma en que debería haberse expresado originalmente en el lenguaje de Tarski, vemos que la estructura de la lista que Field objeta está determinada por el criterio *S*. Pero estos criterios —*T*, o su generalización natural a fórmulas que contengan variables libres, *S*— se hallan determinados por las propiedades formales que queremos que tengan las nociones de verdad y referencia, a causa de que, por diversos fines, *necesitamos* tener un predicado en nuestro metalenguaje que satisfaga precisamente al criterio *S*. (Ésta es la razón por la que *conservaríamos* al criterio *S* aunque eligiéramos una acepción intuicionista o cuasi-intuicionista para las conectivas lógicas.)

Por lo tanto, concluyo que la objeción de Field falla y que es correcto para el realista definir 'verdadero' al estilo de Tarski. A pesar de que la noción de verdad se derive, por decirlo así, mediante una 'deducción trascendental' (el argumento que presenté recién de que *necesitamos* una noción metalinguística capaz de satisfacer al criterio *T*), y se justifique al criterio *S* de un modo parecido, la satisfacción o la referencia continúan siendo, consideradas desde nuestro esquema conceptual realista, relaciones entre palabras y cosas que además poseen valor explicativo, como lo demuestra el razonamiento de Boyd.

Ahora que he desarrollado este argumento, voy a exponer otro, más breve e impreciso, cuyo efecto es más o menos el mismo:

" 'Electrón' se refiere a los electrones", *¿de qué otro modo* podríamos decir a qué se refiere el 'electrón' desde *el interior* de un sistema conceptual en el cual 'electrón' es un término *primitivo*?

Tan pronto como *analizamos los electrones* —a saber, que 'los electrones son partículas con tal y cual masa y unidad de carga negativa'— podemos afirmar que " 'electrón' se refiere a partículas de tal y cual masa y unidad de carga negativa", pero entonces '*carga*' (o cualesquiera que sean las nociones primitivas de nuestra nueva teoría) tendrá que ser explicada 'trivialmente', esto es, de acuerdo con el criterio *S*. Dado el predicamento de Quine (¿de Kant?), según el cual existe un mundo real *pero* sólo podemos describirlo en términos de nuestro propio sistema conceptual (¿deberíamos acaso utilizar el

sistema conceptual de otro?), *¿resulta sorprendente* que la referencia *primitiva* posea este carácter en apariencia trivial? Creo que Field replicaría como sigue: (1) Lo anterior no demuestra que verdad y referencia *deban* ser definidas al estilo de Tarski (es decir, definidas así para un lenguaje escogido y extendidas a otros por medio de traducciones, como ya lo expusimos); y (2) tampoco demuestra que *no pueda* darse una teoría 'fisicalista' de la referencia (o al menos de la referencia primitiva, en el sentido que sea pertinente). Lo único que se ha demostrado es que no se *necesita* una teoría fisicalista de la referencia. Pero pudiera ser (señalaría Field) que ella resultara posible, y que mejorara notablemente nuestra comprensión del fenómeno de la referencia. Después de todo, una teoría fisicalista no sería *incompatible* con las definiciones de verdad/satisfacción de Tarski.

Aún más, Field razonaría que le proporcioné un gran arsenal cuando acepté la explicación del realismo de Boyd. Recién concedimos que referencia y verdad son nociones que participan por lo menos en algunas explicaciones *causales*. Sin embargo, parte importante de la objeción a la analogía de la valencia consistía en que verdad y referencia no son nociones explicativas causales. Y de hecho en cierto sentido *no* lo son; por ejemplo, todavía las necesitamos para construir lógica formal, y también para los otros propósitos ya mencionados, aun cuando las explicaciones de Boyd sobre el éxito de la ciencia resultaran falsas. Pero si participan de alguna manera en las explicaciones causales, ¿no es posible que su función causal explicativa justifique la búsqueda de una descripción fisicalista de lo que *son* la verdad y la referencia, de la misma manera en que la función causal explicativa de 'valencia' justifica la búsqueda de una descripción fisicalista de lo que *es* valencia?

No voy a exponer hoy estas cuestiones; serán tema de mi próxima conferencia.

TERCERA CONFERENCIA

En la conferencia anterior esbozamos la distinción entre entender la verdad y las conectivas lógicas desde el punto de vista realista o "clásico" (es decir, según la teoría clásica de la verdad y la falsedad) y entenderlas de manera cuasi-intuicionista (es decir, en términos de la demostrabilidad en una teoría, digamos B_1).

¿Qué significa entender la verdad y las conectivas lógicas desde la perspectiva realista? Hemos visto lo que no es: el hecho de que se acepte la lógica clásica no demuestra que se las comprenda de manera realista. (Como tampoco el que se rechace la lógica clásica implica una comprensión idealista de las mismas. En otro lugar he propuesto interpretar la mecánica cuántica por medio de una lógica distinta de la convencional, aunque no entendida en términos idealistas, al menos no más de lo que la lógica clásica se debe entender de manera idealista.)[1] No se trata simplemente de aceptar para el propio lenguaje el criterio T o incluso el criterio S, ni una definición de verdad al estilo de Tarski.

La prueba de que interpretamos la noción de verdad desde la óptica realista es que aceptamos enunciados como:

A) Venus puede no tener dióxido de carbono en su atmósfera aun cuando de nuestra teoría se concluye que Venus tiene dióxido de carbono en su atmósfera.

Y

B) Un enunciado puede ser falso aunque derive de nuestra teoría (o de nuestra teoría más el conjunto de oraciones observacionales verdaderas).

[1] 'The Logic of Quantum Mechanics', en mi *Mathematics, Matter and Method; Philosophical Papers*, vol. 1, Cambridge University Press, 1975.

Ahora bien, (*B*) se concluye a partir de cualquier oración que presente la forma general (*A*). Así, ¿por qué creemos (*A*) (y muchas otras oraciones similares)?

Porque, aun si hubiera una ley científica (que por supuesto no la hay) según la cual Venus debe tener dióxido de carbono en su atmósfera, 'Venus tiene dióxido de carbono en su atmósfera' no debería estar *implicado* por 'se concluye de la teoría B_1 que Venus tiene dióxido de carbono en su atmósfera' (donde B_1 es una formalización del conocimiento actual).

Desde luego, 'se concluye de la teoría B_1 que Venus tiene dióxido de carbono en su atmósfera' y B_1 implican *juntos* que 'Venus tiene dióxido de carbono en su atmósfera'; pero la dificultad estriba en que si la afirmación de *p* se reconstruye como la afirmación: '*p* es demostrable en B_1' (o B_1 más el conjunto de todas las oraciones observacionales verdaderas)', *siempre que no haya diferencia* entre 'el acto discursivo' de aseverar *p* y el de aseverar '*p* se concluye de B_1 (más el conjunto de oraciones observacionales verdaderas)', entonces para cada *p*, *p* está *implicado* por '*p* se concluye de B_1 (más el conjunto de oraciones observacionales verdaderas)'. Y así, puesto que cualquier cosa que se siga de B_1 *por sí sola* también se concluye sin duda de la unión de B_1 con el conjunto de oraciones observacionales verdaderas, '*p* se sigue de B_1' *implica p*, para cualquier *p*, lo cual es obviamente absurdo, dadas las preconcepciones realistas normales.

El hecho de que por lo general *rechacemos* la equivalencia entre *p* y '*p* se concluye de B_1 (más el conjunto de oraciones observacionales verdaderas)', muestra que en nuestro concepto actual de verdad resulta *por lo menos lógicamente* posible que un juicio se siga de B_1 (nuestro mejor cuerpo teórico actual) y que no sea *verdadero*. Pero en realidad, creemos que se trata de mucho más que una mera posibilidad *lógica* el que Venus carezca de dióxido de carbono en su atmósfera, pues aun si aceptamos B_1, pensamos que existe la posibilidad de que B_1 sea errónea en un sentido más categórico que el de 'posibilidad lógica' (una 'probabilidad no igual a cero' de que B_1 sea errónea), y *esta* estimación *misma* se encuentra basada en nuestras creencias fácticas acerca del mundo y de cómo lo conocemos. En síntesis, consideramos que el conocimiento mismo es producto de ciertos tipos de interacciones causales, al menos en casos como 'Venus tiene dióxido de carbono en su atmósfera'. Y *de nuestra teoría* de la interacción, por la cual sabemos este hecho (por ejemplo, la

explicación causal común de la percepción y la teoría del error), *se concluye* que podríamos haber elaborado, por la razón que fuera, una teoría de la cual se siga que Venus tiene dióxido de carbono en su atmósfera, aun si no es así. En suma, (*A*) es en sí misma un *dato 'científico' —o incluso del sentido común— sobre el mundo* (aunque se trate de un dato *modal*). Pero, dada la obviedad y la centralidad de nuestro modo de comprender el conocimiento de hechos como (A), ¿puede existir *otra forma* de entender la verdad y las conectivas lógicas que no sea realista? ¿Cómo se podría no ser realista?

Tradicionalmente, la actitud posible radicaba en *aceptar* (*A*), *aceptar* las conectivas lógicas 'realistas' (clásicas), pero dar una explicación idealista del significado de los términos descriptivos (los predicados, o al menos el 'vocabulario teórico'). Pero con el fracaso de los programas reductivos del fenomenalismo y del empirismo lógico, esta vía quedó bloqueada.

Algunos filósofos empiristas —Peirce y más recientemente Sellars— han tratado de preservar la idea de que B_1 puede ser falsa (de hecho, sin duda se volverá falsa) sin recurrir a una noción realista de la verdad y la falsedad, mediante la identificación de la verdad, no con la 'asertabilidad justificada' *actual* (la demostrabilidad en B_1, o en B_1 más el conjunto de oraciones observacionales verdaderas, según nuestra reconstrucción), sino con una asertabilidad justificada *en el límite ideal* de la investigación científica. De este modo, B_1 podría implicar que Venus tiene dióxido de carbono en su atmósfera, pero en el límite ideal de la investigación científica aceptaríamos finalmente una teoría que afirme lo opuesto; y *esto* es lo que Peirce y Sellars consideran que *significa* la aserción de que Venus podría carecer de dióxido de carbono en su atmósfera. Pero lo anterior presupone que podemos darle sentido a la noción de *límite ideal* de la investigación científica sin un marco de locaciones espacio-temporales, objetos, etc., para especificar de qué manera, y en qué condiciones, la investigación científica ha de llegar al límite, lo cual no es el caso. Aún más, esto presupone una convergencia. Si no hay convergencia en general (es decir, si los casos en que ésta no se produce son más numerosos y significativos que aquéllos en que tiene lugar), como piensan Kuhn y Feyerabend, esta teoría de la verdad en el 'límite ideal' es tan deficiente como el realismo.

La actitud más factible, si se supone que el conocimiento científico no converge, sería argumentar que el *fenómeno de las revoluciones científicas* demuestra que la noción realista de referencia (y

por lo tanto de la verdad) conduce al desastre (*a través de* la metainducción que expuse en la conferencia anterior), y así volvemos a caer en la interpretación intuicionista o cuasi-intuicionista de las conectivas lógicas, lo cual salvaría al grueso de la teoría científica extensional y a la parte formal de nuestras teorías de la referencia y la verdad, a costa de abandonar (*A*) y (*B*).

El realista afirma, en efecto, que debe tomarse a la ciencia según su 'valor nominal' —sin reinterpretación filosófica— a la luz del fracaso de todos los programas serios de reinterpretación filosófica de la ciencia, y que la ciencia tomada según su 'valor nominal' *implica* al realismo. (Éste es, por así decirlo, 'la filosofía científica sobre la ciencia'.) Sus oponentes replican, bajo el supuesto de que no hay convergencia, que la ciencia misma considerada *diacrónicamente* refuta al realismo. Pero el fracaso de la convergencia resulta crucial para este tipo de argumento antirrealista. Si estamos en lo correcto al afirmar que las ciencias maduras 'convergen' (en un sentido sumamente complejo) y que tal convergencia posee un enorme valor explicativo para la teoría de la ciencia, este tipo de antirrealismo 'relativista cultural' se desmorona.

En resumen, el realismo depende de la forma de entender la verdad, no sólo de la manera de *definir* la palabra 'verdadero'. El concepto de verdad *no* es filosóficamente neutral. La trampa reside en que el significado de 'verdadero' y de las conectivas lógicas no está fijado por su *lógica formal*. Éstos son términos que encajan muy bien en la explicación del significado presentada por Quine, a saber, que para ellos la distinción entre 'teoría total' y significado de los términos *carece de utilidad*. Sin embargo, según lo expuse, 'teoría total' sólo implica en este caso teoría *total* del conocimiento, y no simplemente teoría *lógica* total, y comprende tanto a nuestra interpretación de la naturaleza como a las interacciones que tenemos con ella. El incluir a (*B*) dentro de la ciencia e impedir la desastrosa metainducción, según la cual 'ningún término teórico tiene referencia', mediante una teoría científica que subraye la relación de los 'casos límite' entre teorías sucesivas, forma parte de una concepción que considera al método científico como *dependiente de nuestro nivel más alto de generalizaciones empíricas sobre el conocimiento mismo*, construido en tanto una interacción con el mundo. Nuestras razones para creer en una compleja versión de la convergencia, como la de Boyd, y para aceptar (*B*) se relacionan con nuestra concepción general del *conocimiento como parte del tema de nuestro conocimiento.*

Los idealistas han sostenido siempre que nuestra noción de verdad depende de nuestra comprensión de la teoría y del acto de 'descubrirla' *como un todo.* Si estoy en lo correcto, ésta es una consideración del idealismo que los realistas necesitan asumir, aunque por supuesto no en la forma planteada por los idealistas.

El nuevo fisicalismo

Nos hemos apartado de nuestro propósito original de presentar una crítica de la explicación 'fisicalista' de la referencia, pero esta digresión nos ha servido porque hemos visto que la elegante réplica de Leeds a Field requiere una interpretación cautelosa. Leeds no se equivoca al sostener que las razones de que necesitemos un predicado en el lenguaje que posea la *propiedad del desentrecomillado* —es decir, de que la adscripción de ese predicado a cualquier oración (en un lenguaje apropiado) equivale a afirmar esta última— no son del tipo explicativo causal y que la teoría formal de la verdad es independiente de consideraciones explicativo-causales. Pero la noción de verdad puede ser usada en explicaciones causales; después de todo, parece posible que el éxito de un comportamiento humano dependa del hecho de que algunas de sus creencias sean *verdaderas.* Y la lógica formal de lo 'verdadero' no es lo único comprendido dentro de la noción de *verdad.*

Con objeto de examinar esta cuestión más profundamente, me permitiré construir un modelo. Quiero imaginar a un hablante que tenga una estructura que, en cierto sentido, sea 'transparente' para nosotros. Para los propósitos de esta historia de ciencia ficción puede tratarse de un autómata que calcula los grados de confirmación (el hecho de que yo no crea que ésta sea una buena manera de trabajar en lógica inductiva no importa en este caso) y las utilidades de acuerdo con una función que lleva integrada, y actúa de tal modo que aumenta al máximo la utilidad estimada. Voy a presuponer que nuestro 'hablante' (llamémoslo Karl) se expresa en un lenguaje 'no natural' —de hecho, según un cálculo funcional de primer orden— y que tenemos las traducciones de las palabras lógicas (que incurren en toda clase de peticiones de principio) pero no del vocabulario extralógico (descriptivo).

Nuestro problema consiste en asignar un 'significado', o al menos una *referencia,* al vocabulario extralógico, justificando por qué

escogimos determinada traducción o definición de verdad según la acepción de Tarski.

Voy a suponer ahora lo siguiente: el hablante es *confiable*, en el sentido de que sus oraciones presentan altas probabilidades de ser verdaderas. (Éste es el caso en el cual resulta más posible que las nociones de verdad y referencia tengan aplicaciones explicativas causales.) Buscaré entonces una traducción o, mejor aún, una definición de verdad, para que su lenguaje justifique esto. Se puede considerar a la condición de que las oraciones del conjunto presentan altas probabilidades de ser verdaderas como una 'restricción' global para la traducción: esto es lo que se denomina 'principio de caridad'. (Bajo esta forma el principio es incorrecto, como lo expliqué en otro lugar, pero me atendré a él como una simplificación deliberada.)

Ahora bien, si existiera una *única* definición de verdad que permitiese al hablante una confiabilidad *máxima* (es decir, si con cualquier otra definición de verdad no equivalente, la probabilidad de que enuncie una oración 'verdadera' fuera menor que la probabilidad de que utilice una oración que resulte verdadera en el sentido de la definición en cuestión) y si siempre fuera razonable elegir esa definición de verdad, la *referencia* quedaría definida de la siguiente manera:

Con el fin de formularla, recordemos que una definición de verdad como la expuesta en la Primera Conferencia determina una relación *referencial* entre los predicados de un lenguaje y los objetos que le son apropiados. Más aún, la relación referencial determina totalmente a la extensión de lo 'verdadero' según se aplica a ese lenguaje. Vamos a presuponer (aunque en realidad no sucede) que existe *una sola relación R* con la cual, si tomamos *R* como la relación referencial de la definición de verdad para el lenguaje de Karl, la probabilidad de que sus oraciones sean verdaderas (en el sentido determinado por la elección de *R* como relación referencial) es mayor que si hubiéramos escogido cualquier otra *R* para ese propósito. En otras palabras, estamos suponiendo que la interpretación de los enunciados de Karl, según la definición de verdad basada en *R*, constituye *el máximo alcance que el principio de caridad puede tener para Karl*. Entonces definamos esto así: Karl *se refiere* a una cosa cuando emplea una palabra si y sólo si acepta esa relación *R* —que conduce a la definición de verdad de mayor beneficio para él— para esa cosa.

Como veremos, esto resulta demasiado simple. Pero detengámonos un momento y observemos lo que ocurriría si *sirviera*:

a) Para calcular la *confiabilidad* de Karl, tenemos que usar no sólo el reconocimiento de su *programa inductivo*, sino también la naturaleza del *ambiente*. Esto, a mi entender, no es inverosímil. En otra parte expliqué que el significado es una función del medio y no únicamente de lo que se encuentra en la cabeza de los hablantes.[2]

b) A qué se refiere Karl depende de la estructura global de su conducta lingüística (e inductiva). Incluso puede decirse que ésta es una característica de *cualquier* teoría de la referencia que incorpore algo *semejante* al principio de caridad.

c) '*X* se refiere a *Y* por la palabra *Z*' *no* constituye una relación *física*, si la definición anterior es correcta: se trata de una relación *funcional*, en el sentido de que no usa predicados que se refieran a la composición físico-química del hablante o del ambiente.

Aun cuando la definición precedente no puede ser correcta por muchas razones, (b) indica que no es posible, por ejemplo, definir a la referencia mediante '*X* se refiere a *Y* si y sólo si *X* (una expresión particular de *X*) se vincula con *Y* por una cadena causal del tipo apropiado'. (Por supuesto, frases como 'cadena causal *del* tipo *apropiado*' son sumamente vagas, así que tal vez las teorías causales pueden eludirlas si permiten que entren restricciones globales para determinar cuál es el 'tipo apropiado' de cadena causal.)

Sin embargo, la definición anterior no sirve. Porque, como Quine lo ha señalado, si una definición de verdad convierte en verdaderas a las expresiones del hablante (o a un subconjunto de ellas), pasaría lo mismo con un número infinito de otras. Es más, aun si pudiéramos fijar de alguna manera cuáles serían las condiciones de verdad para las oraciones del hablante *como totalidades*, el famoso ejemplo del 'gavagai' de Quine, que describiremos más adelante, indica que es posible encontrar definiciones satisfacientes que no sean coextensivas y que conduzcan a condiciones de verdad *equivalentes* para todas las *oraciones* de *L*. De esta forma, 'llevar al máximo la confiabilidad adscrita a Karl', aunque se tratara de una aseveración correcta, podría *subdeterminar* las condiciones de verdad de las oraciones de Karl tomadas como totalidades y la *referencia* de las *partes* de la oración.

[2] "The Meaning of 'Meaning'", en mi *Mind, Language, and Reality; Philosophical Papers*, vol. 2, Cambridge University Press, 1975.

¿Qué estamos haciendo entonces en realidad? Creo que en la traducción efectiva empezamos con supuestos sobre lo que el hablante *quiere* o *pretende*, al menos en muchas situaciones. Después de haber pasado horas sin probar bocado, suponemos que quiere comer; después de permanecer despierto durante horas, quizás quiere dormir (en especial si se restriega los ojos), etc. También suponemos que su 'confiabilidad', en el sentido abstracto de verdad probable, no se vincula con su *eficiencia* personal. Si el hablante acepta una oración *S* cada vez que está mirando el agua, y también cuando siempre que le falta va en su busca, *S* puede significar 'ahí hay agua' o 'ahí hay algo para beber', pero *no es probable* (por decirlo en términos moderados) que *S* signifique 'tres es un número primo'.

Lo anterior sugiere que al 'traducir' construimos una teoría global que explica de un modo razonable la conducta del hablante a la luz de sus creencias (según están determinadas por el manual de traducción que constituye *uno* de los componentes de la teoría global) *y* de sus deseos e intenciones (según se hallan determinados por la teoría psicológica, que es el *otro* componente).

Para evitar una posible malinterpretación, agregaré lo siguiente: no estoy discutiendo solamente que sea una *buena metodología* el descubrir por este medio qué 'significa' el hablante cuando trata de *racionalizar* su conducta. Estoy proponiendo que se debe considerar como traducción o definición de verdad correcta a la que explique mejor la conducta del hablante. Ésta es una teoría metafísica sustancial acerca de qué es 'lo correcto' en lingüística. Al impugnarla, si lo hicieran, los 'fisicalistas' deben ofrecer una teoría alternativa y demostrar por qué es superior.

Ahora bien, racionalizar la conducta del hablante es *explicar*: por ejemplo, estoy dispuesto a *sentarme* en un objeto si y sólo si acepto la oración 'no voy a lastimarme, avergonzarme, etc., si me siento en eso'; y asignar a mi expresión las condiciones de verdad convencionales (suponiendo que por lo general creo en las oraciones que puedo aprobar y que meta, creencia y acción se vinculan de la forma convencional) forma parte de la explicación del hecho de que estoy dispuesto a sentarme en el objeto *Y* si y sólo si acepto el enunciado en cuestión. Este modo de explicar un gran número de hechos acerca de la disposición que el hablante tiene de comportarse de ciertas maneras es precisamente el único sentido que doy a la expresión tan pretenciosa de 'racionalizar la conducta'. Pero la *explicación es una noción de interés relativo.*

Me gustaría ahora desarrollar qué quiero decir con esto y mostrar cómo se interpreta lo que Quine llama 'indeterminación de la traducción', suponiendo que existe, *por el interés relativo de la explicación*.

Interés y explicación

Comenzaré por exponer lo que quiero decir con la 'relatividad del interés' de una explicación y luego intentaré aplicarla al problema de la *traducción*. En realidad, son varios los fenómenos que tengo en mente. Los ilustraré con una serie de ejemplos.

(1) El profesor X es descubierto completamente desnudo en el dormitorio de las alumnas a las 12:00 de la noche. Explicación: (?) estaba completamente desnudo en el dormitorio de las alumnas a medianoche$-\epsilon$, y no pudo salir de allí ni vestirse antes de la medianoche a mayor velocidad que la luz. Pero (ley de cobertura:) nada (en todo caso, no un profesor) puede viajar más rápido que la luz.

(2) Una clavija (de una pulgada cuadrada) embona en un agujero cuadrado de una pulgada y no en un agujero redondo de una pulgada. Explicación: (?) una clavija se compone de determinadas partículas elementales según un cierto arreglo estructural. Calculando todas las trayectorias que se pueden obtener cuando se le aplican diversas fuerzas (con la restricción de que éstas no deben ser tan grandes que deformen a la clavija o a los agujeros) al estilo de la famosa supramente de Laplace, determinamos que hay una trayectoria que hace que la clavija entre en el agujero cuadrado y que ninguna de las trayectorias le permite entrar en el agujero redondo. (Leyes de cobertura: las leyes físicas.)

(3) Supuestamente, se preguntó a Willie Sutton (famoso ladrón): '¿Por qué roba usted bancos?', a lo cual respondió: 'Porque es allí donde está el dinero'. Ahora bien, imaginemos (debo este ejemplo a Alan Garfinkel):
(a) Que es un sacerdote quien formula la pregunta.
(b) Que es un ladrón quien formula la pregunta.

Como dato preliminar, debo decir que considero a la filosofía de la ciencia como una descripción *normativa*. Desde esta perspectiva,

resulta evidente que la explicación tiene que ser en parte un concepto *pragmático*. Pensar que el aspecto 'pragmático' de la explicación no está comprendido dentro del concepto equivale a renunciar a la tarea de entender *qué hace que una explicación sea adecuada*. Más precisamente, el problema no es si vamos a incluir a las características pragmáticas como 'parte del significado', lo cual carece de importancia para el caso de nociones como la de 'explicación'; más bien, se trata de saber si nuestra teoría las justifica o las relega a la simple psicología.

Entonces tenemos que (1) y (2) satisfacen al modelo de explicación 'nomológico deductivo' convencional. Pero son muy deficientes. (La cuestión de si les niega por completo la cualidad de explicaciones —lo cual sería mi concepción de cómo funciona el lenguaje y cómo debería funcionar— o de si sólo se consideran *malas* explicaciones es menos importante.) ¿Por qué?

La explicación (1) es muy deficiente porque toda la deducción se lleva a cabo a partir de premisas que ya sabíamos que eran verdaderas cuando formulamos la pregunta *y* sabíamos que se podía realizar *esta* deducción. Para ahondar en ello: las explicaciones surgen con frecuencia (aunque no siempre) del deseo de encontrar una manera de incluir o eludir a lo explicado, y lograrlo por grados dentro de una categoría específica. Si preguntamos por qué el profesor X se encontraba en una situación tan embarazosa, quizás queramos saber qué motivos, intenciones, neurosis u otras causas *psicológicas* lo condujeron a ello. Incluso *si no estamos interesados* en la prevención o el control, solicitamos el tipo de información que requeriríamos *si ello nos interesara*. Saber que podemos evitar que *X* esté desnudo en el dormitorio a las 12:00 si impedimos que se desnude en ese lugar a las 12– equivale simplemente a devolver la pregunta de una manera poco interesante. Lo mismo ocurriría, incluso, desde el punto de vista del 'entendimiento' puro —una premisa menor dentro de una teoría interesante. (Sospecho que, en realidad, nuestros criterios de 'interés' para el caso del entendimiento puro no se hallan *totalmente* divorciados de nuestros intereses por la predicción y el control, pero no necesito desarrollar esto, lo cual no significa que crea que no valoremos el entendimiento por sí mismo, como en realidad sucede.) De modo que 'las preguntas sobre el porqué', y por lo tanto las explicaciones, *presuponen niveles de interés*.

La explicación (2) es muy deficiente por una razón distinta. No es que *sepamos* o que podamos ofrecer la deducción de la supermente

de Laplace. En realidad, si en principio no se puede dar *ninguna* otra, quizás ésta sería 'la' explicación. Pero por cierto existe otra que sólo recurre al hecho geométrico de que un cuadrado de una pulgada no embona en un círculo de una pulgada y al (los) hecho(s) dado(s) de que la clavija y los agujeros no cambian de forma (y de que lo sólido no pasa a través de lo sólido); esta explicación es más sencilla *y* se generaliza dentro de una clase de casos interesante. De tal modo, (2) transgrede los intereses *metodológicos*, mientras que (1) transgrede los intereses *de la información básica*.

La explicación (3) resulta interesante en otro sentido. El filósofo que concibió este caso, Alan Garfinkel, habla de 'relatividad explicativa':[3] la idea es que una interrogante sobre el porqué de algo siempre presupone un 'espacio' de *alternativas pertinentes*. La pregunta del sacerdote significa: '¿Por qué roba usted bancos —*en lugar de no robar nada*? La pregunta del ladrón significa: '¿Por qué roba usted bancos —en lugar de asaltar, por ejemplo, gasolineras?' Y la respuesta de Sutton es una respuesta (parcial) a la interrogante del ladrón, no a la del sacerdote. (También se podría aplicar la noción de 'espacio explicativo' a (I).)

Retomaré ahora el problema de la 'traducción radical' planteado por Quine. Quisiera modificar el famoso ejemplo del 'gavagai' que este filósofo presenta en el capítulo II de *Word and Object*. De acuerdo con mi versión, tenemos un 'lenguaje de la jungla' con la palabra 'gavagai', a la cual podríamos naturalmente traducir como 'conejo'. De hecho, voy a suponer que, para este lenguaje, contamos con un buen manual de traducción (M), de fácil manejo, en el cual 'gavagai' está efectivamente traducido como 'conejo'. La idea es que *podríamos* elaborar un manual alternativo de traducción (M') en el cual 'gavagai' se traduce como 'parte no separada del conejo' (o 'etapa de desarrollo de un conejo' o cualquier otra cosa). Además, voy a suponer que M' lleva a las condiciones de verdad para *oraciones completas*, las cuales, según se observa de inmediato, son equivalentes a las proporcionadas por M con sólo realizar los ajustes compensatorios en las traducciones de *otras* palabras y/o de los 'taxemas' (características de *ordenamiento* como, por ejemplo, orden de las palabras). (Quine no habla de preservar las condiciones de verdad, sino de lo que denomina 'significados motivadores'.) Ahora bien, la pregunta de Quine es la siguiente: ¿Quién puede decir (y sobre qué

[3] A. Garfinkel, *Explanation and Individuals*, Yale University Press, en prensa.

base) que *M* está en lo *correcto* y M' *se equivoca*, que 'gavagai' significa *conejo* y no *parte inseparada de un conejo*?

Apliquemos el concepto de que *traducir* es *racionalizar la conducta*. Consideremos este caso: el nativo, Karl, ve algo y dice 'gavagai'. Lo mata, lo lleva a su casa y lo cocina. ¿Por qué *a nosotros* nos parece tanto más natural traducir 'gavagai' como *conejo* y no como *parte inseparada de un conejo*?

(1) *No* porque 'simplifique' al manual de traducción *M*. Esto podría ser un factor de peso *al final* (cuando hayamos construido a *M*). Pero creo que *conejo* nos parecería la traducción más natural aun si *gavagai* fuera la *primera* muestra de lenguaje de la jungla que hubiéramos encontrado. (Además, la *razón* de que *M* resulte más simple reside en gran parte en que nuestro lenguaje contiene expresiones cortas para 'conejo' y no para 'parte inseparada de un conejo', 'parte separada del conejo', etc.)

(2) Lo cierto es que encontramos que la *explicación* más simple de la conducta de Karl sería la siguiente: 'Cree que está viendo un conejo. Quiere un conejo para comer. Así que le dispara.' Pero: 'Cree que está viendo una parte no separada del conejo. Desea algunas partes inseparadas del conejo para separarlas y comerlas. Así que dispara a una de las partes inseparadas del conejo en cuanto las ve', nos parece absurdo *dada la manera en que estructuramos el espacio explicativo* y lo que *nosotros* consideramos como clases de casos pertinentes para generalizar, etc.

Pienso que la razón de que la doctrina de Quine sobre la 'indeterminación de la traducción', que en este caso se extiende a la *indeterminación de la referencia* (y, por lo tanto, de las definiciones de satisfacción), parezca tan poco verosímil es que la consideramos desde el punto de vista de *nuestros* intereses, espacios explicativos, etc. Debido a ello, resulta ciertamente poco claro que exista una 'indeterminación'. Bien puede haber una sóla traducción (más o menos) 'correcta' *para nosotros*. Pero examinemos este problema, *no* desde el punto de vista de *una* cultura que traduce el lenguaje de la jungla (y que tiene que escoger entre *M* y *M'*), sino desde la óptica de dos culturas que poseen *intereses diferentes* y que traducen el *mismo* lenguaje. Supongamos que la otra cultura —digamos, la marciana— tiene expresiones cortas para 'partes inseparadas de un conejo y 'parte separada del conejo' (suponiendo, desde luego, que *nuestro* manual de traducción sirve para los marcianos) y que las partes son

de *gran interés* para ella (tal vez porque sus individuos son muy *pequeños* y las partes del conejo, de los árboles, etc., les resultan mucho más evidentes que los árboles y los conejos completos, de los que poseen una concepción menos importante), pero carecen de expresiones cortas para conejos completos porque conejos, gatos, perros, etc., completos, no son de gran interés en la vida cotidiana de los marcianos. Entonces, a ellos bien podría parecerles que la traducción más 'natural' de 'gavagai' es la expresión marciana que *nosotros* traducimos como 'parte inseparada de un conejo'.[4] En suma, la 'indeterminación de la traducción' (y de la referencia) resulta verosímil *en tanto* derive de la relatividad del interés de la explicación.

[4] Un marciano amigo mío ha sugerido que 'tronco de conejo' expresaría mejor la forma marciana de pensar que la pomposa 'parte no separada del conejo'.

CUARTA CONFERENCIA

En las conferencias previas he argumentado que la lógica formal de lo *verdadero* y la *referencia* de una oración es captada por la lógica de Tarski, pero que la misma deja *indeterminados* a los conceptos de verdad y de referencia. De hecho, estas nociones pueden ser pensadas de acuerdo con una definición al estilo de Tarski (para el lenguaje propio), pero es sólo mediante el examen de nuestra teoría del mundo y, específicamente, de las conexiones entre la verdad y los diversos tipos de *demostrabilidad o asertabilidad justificada*, según se esbozan *dentro* de la teoría misma, como se puede determinar si las nociones de verdad y de referencia que empleamos son realistas o idealistas, 'clásicas' o 'intuicionistas'.

Una segunda dimensión de esta subdeterminación aparece, como vimos en la pasada conferencia, cuando extendemos las nociones de verdad y referencia a *otros* lenguajes, a saber, cuando hablamos de lo expresado en otro lenguaje como verdadero o como que se refiere a algo. Esta manera de hablar no sólo presupone la existencia de un concepto de verdad para nuestro propio lenguaje, sino además una proyección del otro lenguaje en el nuestro, y tales proyecciones, como ya lo he señalado, pueden encontrarse relacionadas con nuestros intereses. La 'suavidad' de los hechos sociales podrían afectar la 'dureza' de las nociones de verdad y de referencia. En la presente conferencia deseo examinar algunas posibles objeciones a este punto de vista.

Una de ellas es que:

> interés relativo de la explicación → indeterminación (posible) de la referencia

si se afirma que los intereses del *hablante* son los únicos que importan. *Dados* estos (es decir, suponiendo que *existe* una descripción *psi-*

cológica del hablante, ya sea que podamos *verificarla* en la práctica o no), ¿por qué la referencia y el significado de las palabras del hablante no iban a estar determinados?

Son *dos* los supuestos subyacentes a esta objeción, a los cuales rechazo.

(1) Se presupone (por ejemplo, Peter Winch en su libro *The idea of a Social Science)* que la única explicación aplicable es necesariamente la que el hablante proporciona acerca de su propia conducta. Acepto que siempre *cuenta* (de maneras que son ciertamente difíciles de describir e imposibles de reducir a un algoritmo), pero no siempre es *correcta*.

Para demostrar esto último, señalaré que se debe restringir en cierta medida la relatividad del interés de la explicación. Recordemos que mi punto de vista básico consiste en que la filosofía es (en parte) una descripción *normativa* de nuestras instituciones; la teoría del conocimiento busca explicar y describir las prácticas nuestras que contribuyen al éxito de la investigación. Ahora bien, resulta *útil* para el concepto de explicación el hecho de que rechacemos a las explicaciones que no se adecuan al espacio explicativo, las que sólo se generalizan dentro de una clase de casos 'poco interesantes' (como el ejemplo de la clavija, donde la explicación de Laplace se contrapone a la geométrica) y las que se limitan a respondernos lo que ya sabíamos (el ejemplo del profesor en el dormitorio de las alumnas). Ésta es la razón por la que considero a tales características 'pragmáticas' de la explicación como ilustrativas de un interés *metodológico* y no *meramente* 'psicológico'. Pero de aquí no se deduce que *cualquier* explicación (incluso una 'completa', de acuerdo con el modelo de Hempel) *pueda ser buena si se dan los intereses apropiados*. Porque, debido a nuestra forma *normativa* de ver las cosas, sólo nos interesan las explicaciones que resultan compatibles con ciertos fines de la investigación, como *formulación de verdades, descubrimiento de errores, postulación de leyes* (para el caso de la investigación científica), *establecimiento de lo que es posible que ocurra y lo que es imposible, descripción de diversas clases de procesos importantes, contribución general a la tecnología y al control de la naturaleza, etc.* Interesarse por la razón de que Sutton robe bancos en lugar de estaciones de gasolina y de que robe bancos en vez de no robar nada es, *en ambos casos*, compatible con estos objetivos. Por ello éste representa un ejemplo genuino de la relatividad del interés.

Pasando al ejemplo de los 'marcianos', aun cuando éstos explicaran la conducta de Karl como que manipula 'partes no separadas del conejo' en lugar de conejos, y a pesar de que, dados sus intereses, esquema conceptual, etc., ello resultara 'natural' para los marcianos, para que se convierta en un caso *interesante* todavía tenemos que suponer que sus intereses no son arbitrarios (proponer, digamos, que son de tamaño pequeño); y además que son compatibles con el desarrollo de una ciencia que contenga un conjunto de 'leyes' tan buenas como las nuestras, que sean por lo menos tan aproximadamente verdaderas como las nuestras, que posean eficacia tecnológica, etc. Y este ejemplo tiene interés precisamente a causa de que no hay razones para creer que esto no sea así —una cultura que hable de las 'partes del conejo' *podría* comportarse de este modo. Por cierto, cabe notar que, cuando lo desean, los marcianos *pueden* referirse a los conejos como 'totalidades de conexión máxima que se componen de partes no separadas de conejos', y que pueden darse cuenta perfectamente de que los conejos son más *importantes* que sus partes en *ciertos contextos*, por ejemplo, dentro de la teoría biológica. Mi suposición se limita a que en su *vida cotidiana* les resulta más natural pensar y hablar de 'haber visto partes no separadas de un conejo', etc., que de 'haber visto conejos', etc.

Ahora bien, es indiscutible que cuando nosotros *o* los marcianos explicamos la conducta de Karl de acuerdo con nuestros modelos respectivos, no sólo tenemos que considerar lo que hace en el contexto inmediato (dice *gavagai*, mata al conejo, lo cocina y lo come), sino también su conducta total *según como la racionalizamos* (por *traducción*); en particular, la descripción que nos proporciona acerca de sus propios *intereses, intenciones y deseos*. Pero eso *no* significa que Karl deba aceptar nuestra explicación final de su conducta total, dados sus intereses, etc.[1]

[1] Tampoco se trata necesariamente de que el nativo (por ejemplo Karl) nos pueda ayudar si así lo quiere. Después de todo, el 'hablante nativo' de la teoría de los números no puede 'decirnos' a qué número *se refieren* las palabras (como se quejaba Frege); *nosotros* tenemos que decidir desde la posición superior de *nuestro* lenguaje. (Supongamos que *sólo* hemos determinado el habla y no un vocabulario numérico, y que por primera vez nos topamos con una cultura que habla de 'números'.) ¿Por qué no podría suceder que Karl, después de aprender *tanto* el inglés *como* el idioma de Marte, nos dijera: 'si me fuerzan a pensar en el problema, me parece que el término '*gavagai*' es, en cierto sentido, *ambiguo*: puedo *construirlo* como conejo o como parte no separada del conejo. Pero el idioma gavagués no es ambiguo en sí mismo; yo sólo pienso en los gavagais como gavagais'?

Como ejemplo, supongamos que Karl pertenece a una cultura basada en una teoría de los datos de los sentidos. Karl puede afirmar que la mejor explicación de por qué dice *gavagai* consiste en que 'percibe directamente los datos sensoriales que produce un conejo'. Y quizás rechace las explicaciones que hablan de que 've a un conejo' como 'metafísicas'. Pero esto no implica que *nosotros* tengamos que hallarnos necesariamente de acuerdo con que ésta sea una buena explicación de la conducta verbal de Karl, ni que *gavagai* signifique 'mira, ahí va un dato sensorial producido por un conejo', aun si Karl *afirma* que eso es lo que él quiere decir. Los enunciados metalingüísticos explícitos pueden contener todo tipo de teorías falsas.

El problema consiste en que la 'relatividad del interés', una vez concedida, no puede ser evitada tomando las interpretaciones de *Karl* como las correctas, según pretende Winch.

(2) El segundo presupuesto que subyace a esta objeción consiste en que hay una descripción de los intereses de *Karl* que es objetivamente correcta y *no es relativa a intereses* (es decir, los *nuestros*), —una verdadera 'descripción psicológica' de Karl—, que podemos o no verificar. Quiero refutar esto. En aras de la precisión, supondremos que los seres humanos son máquinas de Turing[2] (lo que *por cierto* no creo). Y concederemos (temporalmente, al menos) que existe una descripción psicológica de Karl que es correcta en términos 'objetivos' *según el diagrama de la máquina*. Observemos que esto es *del todo* diferente a suponer que existe una descripción objetivamente correcta de Karl *en el sentido de una teoría psicológica*, si por ello entendemos una teoría que emplea nociones como *interés, deseo, creencia*, etc. Mi principal argumento será el siguiente: la indeterminación de la traducción es equivalente a la indeterminación que se observa al pasar de la organización funcional (en el sentido del diagrama de la máquina o cualquier otro) *a* la descripción psicológica.

[2] Una Máquina Turing es un autómata de estados finitos con una memoria potencialmente infinita (bajo la forma de una cinta de papel bidireccional potencialmente infinita en la que la máquina puede imprimir o borrar, y mover a derecha o izquierda). El uso que se da a los modelos autómatas en la filosofía de la mente se discute en algunos de mis artículos de *Mind, Language and Reality; Philosophical Papers*, vol. 2, Cambridge University Press, 1975. Para una explicación general, véanse particularmente los artículos de ese volumen titulados 'Tha Nature of Mental States' y 'Philosophy and our Mental Life'.

El problema reside en que el diagrama de la máquina (o cualquier generalización realizada a partir del mismo) no 'clasifica' a los estados o configuraciones *en términos psicológicos*. Incluso si poseemos el diagrama, todavía tenemos que deducir la configuración que corresponde a *creer que P*, etc. El hecho de que el marciano no coincida con nosotros en cuanto a la traducción de las expresiones de Karl implica que no estará de acuerdo con nosotros (en realidad, no *puede* estarlo si se atiene a su modelo de traducción) en lo tocante a 'clasificar' las configuraciones *funcionales* con términos psicológicos. Supongamos que ambos aceptamos que la configuración mecánica de Karl es *C* cuando éste dice *gavagai*. Pero el marciano afirma que el 'término' psicológico apropiado para la configuración *C* es 'creer que se manifiesta una parte no separada del conejo', mientras que *nosotros* afirmamos que es 'creer que se manifiesta un *conejo*'. Aquello que hace que para el marciano un término resulte más natural que el otro también hace que una traducción de *gavagai* le parezca más natural, y viceversa. De modo que ser 'realista' con respecto a la psicología de Karl *en el sentido de su organización funcional* no es de gran ayuda.

¿Pero por qué limitarnos a ser realistas en lo tocante a la organización funcional? ¿Por qué no serlo únicamente con respecto a las intenciones, los deseos, las creencias, en suma, a la descripción psicológica en el sentido ordinario?

En parte, la respuesta reside en que esto facilita demasiado contestar a Quine. Lo que un hablante significa cuando dice *U* podría o puede determinarse observando cuál es la *creencia* que normalmente acompaña a *U* (o tal vez cuál es la que uno *trata* de transmitir al enunciar *U*, al estilo de Grice[3] y Schiffer).[4] Esto se parece a lo que Quine denomina el 'mito del museo' para significados vestidos con trajes de gala psicológicos. (El 'mito del museo' es la inútil teoría de que existen objetos llamados 'significados', y lo que una oración quiere decir se halla determinado por los objetos de este tipo a los cuales se encuentra 'ligada'.) Pero quizás el mito del museo es *verdadero*. Tal vez resulta fácil en efecto responder a Quine. (Así parece creerlo Noam Chomsky.)

[3] *Cfr.* su artículo 'Meaning', *Philosophical Review*, vol. LXVI, núm. 3, 1957, pp. 377–386.

[4] *Cfr.* su libro *Meaning*, Nueva York, Oxford University Press, 1972.

Realismo y descripciones equivalentes

Consideremos ahora la siguiente información acerca de Karl. Supongamos que tenemos (1) la organización funcional de Karl (el diagrama de la máquina o algo semejante); (2) las 'clasificaciones' psicológicas (desde *nuestra* perspectiva, *nuestro* esquema de traducción o lo que sea) de las diversas configuraciones. Me referiré a esta última como la *teoría* psicológica de Karl. De hecho, se trata de una teoría psicológica muy *completa.* Dado que muchas configuraciones pueden no tener 'nombres psicológicos' —puede haber muchos tipos de 'variables en juego'—, ésta proporciona una información mucho más amplia acerca de los procesos internos de Karl (sus procesos 'inconscientes', en cierto sentido) que cualquier otro conjunto de enunciados que *sólo* utilicen un lenguaje 'psicológico' en su acepción ordinaria. Llamemos *T* a esta información sobre Karl. Supongamos que *T'* es igual a *T*, excepto en que usamos los términos psicológicos del marciano al construirlo. De tal modo, *T* y *T'* estarán en estrecha relación; de hecho, habrá un 'esquema de traducción' capaz de convertir a cualquier descripción del lenguaje *T* en una descripción del lenguaje *T'* y viceversa. *T* y *T'* poseen *la misma clase de modelos.* Según un criterio razonable, constituyen *descripciones equivalentes* de la psicología de Karl, aun cuando uno diga 'Karl piensa que ve un conejo' y el otro 'Karl piensa que ve una parte no separada del conejo'.

Ahora bien, resulta fácil percibir el error del 'realismo' psicológico de Chomsky: *ignora la existencia de ese par de descripciones equivalentes.*

Un realista del siglo XX *no puede* pasar por alto la existencia de descripciones equivalentes: el realismo no está obligado a presentar una sola (y *única*) teoría verdadera. Esto es tan cierto para la psicología como para cualquier otra disciplina. Resulta ingenuo suponer que hay un 'hecho positivo' que nos dirá 'cuál es la verdadera' (si alguna lo es) *siempre* que tengamos dos teorías 'diferentes' desde el punto de vista intuitivo. Las 'teorías' que difieren con respecto a la *simultaneidad* de un par de hechos son, sin duda, 'distintas desde el punto de vista intuitivo', pero desde Einstein sabemos que tales 'teorías' pueden ser, al menos, *equivalentes.* Si el ejemplo del *gavagai* pudiera ser desarrollado (si esos 'marcianos' *existieran*), proporcionaría un ejemplo parecido de 'teorías diferentes desde el punto de vista intuitivo', aunque *equivalentes*, para la *psicología*: ésta es la

razón de que preguntar '¿por qué no ser simplemente tan realistas en psicología como en física?' no es *responder* a Quine. La posición contraria consiste, desde luego, en que *debemos* ser 'realistas' en ambos campos, realistas *refinados*. Y esta clase de realistas reconoce la existencia de varias descripciones equivalentes, porque de su teoría del mundo se concluye que las hay, de igual modo que de la descripción de la tierra realizada por un geógrafo se concluye que hay mapas alternativos (mercator, polar, etc.).

Digresión

Hay otro argumento posible al que no pretendo detallar (porque no lo conozco con seguridad y es demasiado escabroso) pero que deseo mencionar. A saber, se puede afirmar que la organización funcional es *en sí misma* una noción de interés relativo. Consideremos el caso más simple, la aserción de que alguien es una máquina de Turing de un cierto tipo (especificado mediante la gráfica T para la máquina). Digamos que T es una gráfica de estado doble. Ahora bien, algo que nadie *tomaría* por una máquina del tipo T puede resultar que lo es efectivamente mediante una forma extravangante ('artificial') de especificar cuales son los estados S_1 y S_2. ¿En qué consiste esta 'extravagancia' o 'artificialidad'? No es, como se podría suponer en principio, que en un ejemplo 'kosher' sobre una máquina del tipo T, los predicados microestructurales P_1 y P_2 que 'llevan a la práctica' a los 'estados lógicos' S_1 y S_2, *definidos en abstracto*, deban ser 'predicados naturales' desde el punto de vista de los físicos. Ello descartaría a predicados como 'ser una inscripción de la letra cursiva *a*', que parece enteramente correcto. Más bien se trata de que la gráfica (al colocar P_1, P_2 en el lugar de los 'estados lógicos' S_1, S_2 a los que 'llevan a la práctica') ha *explicado* de un modo conveniente el comportamiento de la máquina. De tal manera, podría afirmarse que la noción de organización funcional también incluye una referencia encubierta a la *explicación*, y en consecuencia puede 'heredar' la relatividad del interés de esta última.

La razón de que este argumento no me parezca tan convincente como el anterior es la siguiente:

Si una *simple* diferencia en los intereses provoca que la gente considere a diferentes predicados como 'naturales' (*conejo en contraposición a partes de un conejo*), ambos conjuntos de intereses serían compatibles en líneas generales con los intereses de la ciencia. Esto

no se justificaría si todas las explicaciones y descripciones formaran *parte* de la ciencia teórica y experimental, o si el avance de esta última fuera el objetivo de toda explicación. Pero cuando explicamos la conducta de una persona diciendo, por ejemplo, 'se enojó', no nos preocupamos por el hecho de que 'enojado' sea un término bastante amplio del lenguaje ordinario, y no un término perteneciente a la teoría psicológica; ni por el hecho de que, sin duda, una teoría psicológica ideal sustituiría la noción de 'enojado' correspondiente al lenguaje ordinario por una legión de conceptos técnicos, de la misma forma en que la física remplaza la noción de energía del lenguaje ordinario por un conjunto de conceptos técnicos (algunos de los cuales conservan su antiguo nombre, por ejemplo, la energía cinética). Aunque las buenas explicaciones no deben obstaculizar el progreso de la investigación científica, ello no implica que las explicaciones concebidas para los propósitos de la vida cotidiana, de la industria o incluso de la ciencia aplicada, tengan que usar las categorías que serían ideales desde el punto de vista de la ciencia pura. Los marcianos pueden describir perfectamente la situación que nosotros expondríamos como 'cazar un conejo para cocinarlo y comerlo', diciendo que se trata de 'cazar partes no separadas del conejo para cocinarlas, separarlas y comerlas'. ¿No son éstas en realidad dos descripciones de la misma intención? Sin embargo, la aplicación que den a las 'partes del conejo en las conversaciones cotidianas no les impide poseer un término *técnico* para 'conejo', que pueden utilizar en los contextos científicos apropiados.

Hasta cierto grado, los intereses pueden provocar que se contemplen los procesos del mundo como si exhibieran causas diversas y, no obstante, siguieran siendo igualmente compatibles con la ciencia. Para usar un ejemplo que, en esencia, debo a Hart y Honore:[5] el médico puede considerar que un ataque al corazón fue causado por la alta presión arterial, mientras que la esposa del paciente quizás piensa que su causa radicó en una excesiva indulgencia en el comer y el beber; y ambos pueden tener razón. Pero éste es un problema que se deriva de centrar la atención en distintas partes de un mecanismo causal, mientras que la diferencia con respecto a la gráfica de la máquina se refiere al mecanismo causal *global*. Me parece una falsedad que los intereses puedan conducirnos a creer

[5] H. L. A. Hart y A. M. Honore, *Causation in the Law*, Oxford University Press, 1959.

que los procesos del mundo exhiben mecanismos causales globales diversos, pero que siguen siendo igualmente compatibles con los intereses de la ciencia. Así, la relatividad del interés (en principio) de una organización funcional puede carecer de todo atractivo, mientras que la correspondiente al pasaje *desde* la organización funcional hacia la psicología resulta tal vez en extremo interesante. Puesto que teorías psicológicas intuitivamente diferentes pueden coincidir con la *misma* organización funcional, son capaces de, por ejemplo, vincularse igualmente bien con la biología molecular, etc. (Como todas las descripciones equivalentes, las T y T' ya citadas exhiben *invariantes*, su organización funcional.)

Otra cuestión que no termino de entender es la siguiente: *si* se descubriera que la organización funcional posee un interés relativo importante, ¿nos *contentaríamos* con esto o buscaríamos *suprimir* la relatividad *restringiendo* la clase de las P que pudieran ser realizaciones admisibles de los estados lógicos S? (Una de las razones por las que tal vez deseáramos hacer esto reside en que de otro modo, si el funcionalismo no se equivoca, predicados como *consciente* resultarían ser de interés relativo.)

Otra objeción

Ésta es bastante vaga, pero me la han planteado a menudo, especialmente los seguidores de Chomsky. Por lo regular, se expresa de esta manera: 'En tanto una explicación es de interés relativo, lo es para todas la ciencias, para la física al igual que para la psicología. Por lo tanto, no se ha demostrado que exista alguna razón especial para pensar que la psicología (y la traducción) presente una indeterminación mayor que la *física*'.

En realidad, lo anterior parece ser una mala respuesta a Quine, porque éste argumenta que se observa una 'determinación insuficiente' tanto en física como en psicología *por las mismas razones*, y la 'determinación insuficiente' en psicología *persiste* incluso si 'precisamos' a la física.

Mi respuesta es ésta: no pretendo que la relatividad del interés de la explicación implique *siempre* que podemos obtener *descripciones equivalentes* en cada campo y en cada vocabulario. Es más frecuente que lleve a teorías que *contestan diferentes preguntas*. En el ejemplo del *gavagai*, según las modificaciones que le hice, lo que probablemente sucedería al principio es en realidad lo siguiente:

el marciano se preguntaría por qué Karl ha señalado las partes no separadas de un ganso, y no las de un hombre, etc.; *nosotros* nos preguntaríamos por qué señaló a un conejo y no a un ganso, un árbol, un hombre, etc. Pero si ambos esquemas de traducción deben ser compatibles con alguna psicología científicamente aceptable, tendrán que serlo con teorías psicológicas que constituyan descripciones equivalentes, aun si son 'diferentes' desde el punto de vista intuitivo. Repito que *no hay* un principio metodológico que diga que éstas siempre existen. Por otra parte, aun cuando *tenemos* descripciones equivalentes en, digamos, la física, no siempre figuran en el mismo vocabulario. La teoría de la acción de las partículas a distancia y la teoría de la acción de los campos por contacto pueden ser descripciones equivalentes, pero no emplean el mismo vocabulario.

Para demostrar que T_1 y T_2 son descripciones equivalentes (en el mismo vocabulario), comúnmente se prueba que coinciden con ciertas 'invariantes' *y* se argumenta (según el papel explicativo de las invariantes) que esto es lo único en lo que deben coincidir: que las invariantes proporcionan un cuadro 'completo' de lo que está sucediendo. Además, T_1 y T_2, o mejor dicho la 'equivalencia' de T_1 y T_2, no debe ser un caso de lo que ha dado en llamarse 'convencionalidad semántica trivial', con un simple cambio arbitrario de la referencia de una palabra. Es porque Einstein pudo argumentar que (1) *sus* invariantes proporcionan efectivamente una descripción física completa de los procesos, y (2) emplear 'simultáneo' para significar 'simultáneo en el Marco 1' es tan correcto (de acuerdo con la costumbre o lo que sea) como utilizarlo para implicar 'simultáneo en el Marco 2', que la relatividad de lo simultáneo en la Teoría Especial de la Relatividad representa un ejemplo *no trivial* y *correcto* de descripciones equivalentes.

En esta conferencia he intentado sugerir un conjunto de *invariantes psicológicas* —la organización funcional— de tal naturaleza que los diferentes esquemas de traducción puedan *concordar* con ellas de un modo verosímil. Esto constituye al menos el esbozo de un modelo *específico* en el cual la traducción puede adquirir un interés relativo: sin deducir que debe tenerlo por algún principio metodológico universal. Debido a que las explicaciones psicológicas del lenguaje ordinario no se encuentran tan restringidas a nivel lógico como las especificaciones de una organización funcional, pueden ser permu-

tadas *dentro* de los límites fijados por una especificación (hasta el isomorfismo de los modelos) de la organización funcional. Y dado que esta última constituye (según creo) una noción apropiada acerca de la 'descripción invariante' en *psicología*, lo anterior significa que las teorías psicológicas del lenguaje ordinario y los esquemas de traducción pueden ser *distintos* pero equivalentes.

QUINTA CONFERENCIA

La relatividad del interés de la explicación puede facilitar la traducción en un principio (si Karl dice *'gavagai'* y la idea es siempre un conejo, lo traducimos como *conejo* sin preocuparnos acerca de posibilidades como 'calidad de ser conejo otra vez' o 'partes no separadas del conejo'); pero si lo que hacemos es *forzar* nuestro patrón de pensamiento sobre un esquema lingüístico que no posee similitudes *estructurales* con el nuestro, un esquema que no 'coincide' con el nuestro bajo ninguna proyección 'natural' (para nosotros), es probable que debamos pagar el precio más tarde, cuando intentemos extender nuestras primeras traducciones de expresiones aisladas a un esquema capaz de interpretar el lenguaje total.

Naturalmente, lo anterior suscita la pregunta de que, si la traducción (como todas las explicaciones psicológicas) se relaciona en principio con el interés, por qué no es más difícil de lo que parece.

La respuesta, en parte, consiste en que *lo es*. Como hijo de un traductor, sé perfectamente que la tarea no es fácil. Y la simple observación de traducciones diferentes ilustra en cierta medida la relatividad del interés que se verifica en la práctica. Sin embargo, cuando se trata de palabras comunes y locuciones poco refinadas, la traducción es bastante determinada. Nadie que sepa hablar hebreo moderno, por ejemplo, duda que *galgal* significa rueda y no 'parte no separada de la rueda'. ¿Cuál es el porqué de tal determinación?

Para volver a lo nuestro, señalaré que la 'traducción radical' constituye una *habilidad enseñable*. Desde hace años, Kenneth Pike, profesor de lingüística de Michigan, *ha demostrado en la práctica* que más o menos en una hora puede aprender lo suficiente acerca de un lenguaje totalmente extraño como para mantener una conversación. Y *enseña a sus estudiantes* a hacer lo mismo (algunos de ellos irán luego al Amazonas, donde esta habilidad puede salvar sus vidas).

Si la traducción fuera tan indeterminada en la *práctica* como pudiera resultar en sus principios, no habría posibilidad alguna de aprender la 'habilidad' de Pike (de hecho, él carecería de ella) y no existiría la *capacidad universal de intercomunicación entre las culturas humanas*.

Sólo parece haber una explicación posible: los *intereses*, los *detalles relevantes* y los procesos cognoscitivos de los seres humanos deben presentar una *estructura* fuertemente determinada por factores innatos o constitutivos. La naturaleza humana carece de *semejante* plasticidad.

Pero de ser así, se plantearía la pregunta de por qué no tomamos los intereses, los detalles relevantes, las 'restricciones' que los seres humanos introducen de manera innata a la traducción (como opuestos a los que constituyen meros artificios culturales) y los 'incorporamos' a la *noción* de traducción. ¿Por qué no superamos la 'relatividad del interés' mediante la estipulación de aquéllos, considerándolos en esta ocasión como intereses, puntos relevantes, prioridades, etc., *invariantes* y especificados biológicamente? Por supuesto, no necesitan ser enteramente innatos, puesto que bastaría con que lo innato fuera una estructura que garantizara cierta *propensión* por determinados intereses dentro de una categoría muy amplia de ambientes (los cuales incluirían a todos o casi todos los medios humanos físicos y culturales efectivos).

Pienso que esta propuesta contiene tres elementos erróneos.

Aun cuando pudiéramos determinar 'las restricciones de la traducción radical' que empleamos de manera intuitiva y preconsciente (si no *in*consciente), 'incorporarlas' a la *noción* de traducción sería un error. Pues:

(1) Si este programa tuviera éxito, su resultado consistiría en un análisis operativo de la traducción ridículamente complicado (de hecho, una 'definición operativa'), en lugar de una teoría sobre su *naturaleza*. Aunque lo segundo es lo que una buena explicación de la traducción debería darnos. La teoría de que una traducción a nuestro lenguaje es una correlación que usamos para proyectar nuestros patrones de explicación psicológica —los que aplicamos a los hablantes de *nuestro* lenguaje— o una aproximación a dichos patrones, en los hablantes de un lenguaje extraño, *proporciona* en efecto una teoría sobre la naturaleza de la traducción, aun cuando no produce 'restricciones' en una forma obvia.

(2) Al igual que todos los análisis operacionales, éste resultaría *erróneo* (salvo como aproximación). Una definición de 'traducción' de la cual se siguiera que nuestros hipotéticos marcianos no estaban *traduciendo* (porque sus intereses y sus rasgos característicos son diferentes) sería simplemente incorrecta.

(3) Dicho programa (el que explica la noción de traducción a través de un conjunto de 'restricciones', ignorando por el momento la procedencia de estas últimas, ya se trate de la naturaleza humana innata, nuestra cultura o lo que sea) es completamente inviable en la *práctica*.

Se invirtieron varios millones de dólares y muchas horas de trabajo de gente brillante *únicamente* en este proyecto —explicitar las restricciones implicadas en la traducción— bajo el título de 'traducción mecánica'. Su resultado: en esencia, *un completo fracaso*. Chomsky dijo desde el principio que no se podría simular una traducción si no se simulaba también la totalidad de la capacidad humana, y se demostró que tenía razón. Marvin Minsky (figura destacada del Equipo para el Estudio de la Inteligencia Artificial del MIT) ha presentado recientemente un punto de vista parecido bajo el título de 'teoría de los marcos referenciales'. Su idea es que no se puede traducir un discurso si no se lo coloca dentro de un 'marco' de información pertinente que esté organizado de una manera apropiada. Desde la perspectiva de Minsky, no existe una teoría de la traducción *en general*, sino sólo una teoría para cada 'marco de referencia'. A fin de simular la traducción *en general*, o la traducción radical, habría que contar con una simulación de toda la inteligencia humana, sus rasgos característicos, etc.

¿Cuál es la moraleja de esto? La que me interesa plantear es más o menos como sigue: *la noción de traducción* (y si mi razonamiento es correcto, todo lo que estoy diciendo acerca de la traducción sirve también para la *referencia*) *no puede ser convertida en algo científicamente preciso* (por lo menos no en el futuro cercano y muy probablemente nunca), *pero resulta perfectamente aplicable en la vida cotidiana e incluso en partes de la ciencia lingüística* (los diccionarios, por ejemplo, forman parte de esta última). (Me parece de suma importancia el hecho de que un término, que no puede ser convertido en algo 'científicamente preciso', sea quizás indispensable *aun para la ciencia*, y me extenderé sobre el particular en la siguiente conferencia.)

Ahora bien, el programa de *definir* a la referencia de acuerdo con las 'cadenas causales', etc. (el programa de Field, con el que empecé estas conferencias), puede parecer diferente al de *definir* la traducción mediante un conjunto de 'restricciones', pero la diferencia es superficial. Tan pronto como se intenta ampliar la teoría causal para que comprenda, por ejemplo, términos científicos teóricos ('electrón'), entra en juego, en una u otra versión, el principio del beneficio de la duda, el principio que dice que para saber a qué se refería Bohr con 'electrón' en 1904, debemos ver cuáles son, entre las reformulaciones razonables de las descripciones que aquél presentó, las que no refieren (porque, por ejemplo, violan el principio de complementaridad que Bohr mismo enunció treinta años después); pero ofrecer un análisis *preciso* de la noción de reformulación *razonable* de una descripción determinada es, si acaso, *más* inútil que elaborar una lista precisa de las 'restricciones' de la traducción. Y los problemas son muy parecidos: tanto la asignación de la referencia como la traducción dependen de que se escoja 'razonablemente' dos conjuntos de creencias que no coinciden con exactitud para hacerlos concordar. Simular (o más precisamente definir) lo 'razonable' equivale a simular (o por lo menos definir) la capacidad humana total. En suma, el programa de Field es una especie de *utopismo científico*.

Con lo anterior no estoy repudiando a la 'teoría causal de la referencia' (a la que yo llamaría más bien 'la cooperación social sumada a la contribución de la teoría ambientalista sobre la *especificación* de la referencia') presentada por Kripke[1] y que yo desarrollé en "The Meaning of 'Meaning' ".[2] Kripke y yo hicimos dos cosas:

1 Atacamos la idea de que los hablantes escogen sus referentes de la siguiente manera: 'asocian' cada término T con una propiedad P_T (el 'contenido' de T). El término se aplica a cualquier cosa que tenga la propiedad P_T.

2 Proporcionamos una explicación alternativa sobre el modo en que los hablantes escogen *efectivamente* sus referentes si no asocian los términos con condiciones necesarias y suficientes (o propiedades P_T), como lo exige, por ejemplo, la teoría de Russell.

[1] Los puntos de vista de Kripke se expusieron en unas charlas celebradas en Princeton y se reimprimieron bajo el título de 'Naming and Necessity' en D. Davidson y G. Harman (eds.), *Semantics of Natural Language*, Dordrecht, Reidel, 1972.

[2] *Cfr.* 'Explanation and Reference' y 'The Meaning of "Meaning" ', reimpresos en mi *Mind, Language and Reality; Philosophical Papers*, vol. 2, Cambridge University Press, 1975.

Ambos puntos me siguen pareciendo correctos y de gran provecho. Pero una teoría sobre cómo *se especifica* la referencia no es una teoría acerca de *qué es* esta última; de hecho, *presupone* a la noción de referencia.

Retomaré brevemente el problema de la traducción mecánica. En realidad, este proyecto no consistía en 'hacer explícitas las restricciones de la traducción', según ya lo mencioné. Se trataba de explicitar lo que Quine llama 'hipótesis analítica' o 'manual de traducción'. Pero ello sirve incluso mejor a mi razonamiento: las 'restricciones de la traducción' son reglas para escoger entre distintos manuales de traducción, y resulta que estos últimos *existen* sólo y únicamente como una aptitud humana organizada en un cuerpo.

Quiero dedicar lo que resta de esta conferencia a analizar con mayor detenimiento el 'utopismo científico'. Esto allanará el camino para la conferencia siguiente (y última), la cual tratará el carácter de las ciencias (o estudios) sociales en general.

¿Acaso estoy diciendo que no podemos hacer algo 'en principio'?

Siempre que se utiliza un calificativo como 'utopismo científico', cabe esperar ciertas preguntas. Tarde o temprano, y generalmente más temprano que tarde, se nos interrogará si queremos decir que *no se pueden* conocer 'en principio' las leyes que gobiernan a la psicología humana, o lo que sea, si estamos afirmando que hay algo que podemos hacer en física pero que es imposible de realizar 'en principio' en psicología. Quiero argumentar que, con respecto a un significado importante del término 'en principio' —aunque no se trate de una acepción en la que el cuestionador, en su entusiasmo por el fisicalismo, se detenga a reflexionar — la respuesta puede ser 'sí'. Pero aún más relevante, en mi opinión, es el hecho de que esto equivale en gran medida a plantear la pregunta equivocada, a menos que vaya acompañada por *algún* razonamiento acerca de si lo que puede hacerse 'en principio' es también realizable en la *práctica*.

Obviamente hemos progresado en nuestra comprensión acerca del hombre y la sociedad. Aunque son pocos los sociólogos 'marxistas', en el sentido de que sostienen que el trabajo de Marx es inmune a la crítica o de que comparten la creencia marxiana en un futuro utópico, algo que Marx enseñó a todos los estudiosos de la sociedad es que las formas sociales se suceden entre sí, un poco como se suceden entre sí las especies de árboles en un bosque, de modo que una

provee las condiciones favorables para la aparición de la siguiente; y que la organización económica impone importantes restricciones a la organización social, si bien resulta altamente controvertido cuáles son esas limitaciones y cómo operan. Y a pesar de que son escasos los psicólogos 'freudianos', algo que Freud nos enseñó a todos es que lo que llamamos 'vida consciente' emerge y flota en la superficie de un vasto océano de 'fantasías egoístas' (para usar la terminología de Iris Murdoch), las cuales son inconscientes en su mayor parte. En el otro extremo —más referido a las matemáticas— de la ciencia social, los modelos cuantitativos han arrojado luz sobre fenómenos que van desde la macroeconomía hasta lo que ha dado en denominarse 'ley del efecto' en la teoría del aprendizaje. Pero es importante advertir que cuando uno sueña con definir explícitamente una noción como 'traducción', 'explicación' o 'razonable', está soñando con algo completamente diferente a todo lo que se ha realizado hasta ahora. Como ya mencioné, *todo lo que conocemos* se utiliza en la traducción (al igual que en la explicación y cuando se decide qué es lo 'razonable'). No existen motivos para creer que se puede definir explícitamente alguna de estas nociones de acuerdo con la forma deseada por los filósofos fisicalistas, ni explicitar todas las 'restricciones' que imponemos de manera inconsciente al uso de esas nociones, sin contar con *un detallado modelo explicativo sobre la organización funcional del ser humano total*. Y afirmo que esperar esto es utópico.

Modelos explicativos y clases naturales

Dejaré nuestro tema momentáneamente para exponer qué significa tener un detallado modelo explicativo de una clase natural. Nuestro paradigma para ese modelo consiste en un conjunto de leyes acompañadas por una descripción de las características que presenta un miembro de la clase natural, de tal forma que nuestra supermente laplaciana (en sus conversaciones informales, Carnap solía llamarla 'el fulano lógicamente omnisciente') pueda deducir 'en principio' el comportamiento (o la distribución estadística de conductas, en el caso de la teoría indeterminista) de un miembro de la clase natural a partir de las leyes y de (a) la descripción en cuestión, (b) los valores de diversos parámetros para el caso del miembro de que se trate, (c) los valores de diversos parámetros que describan las 'condiciones iniciales' y las 'condiciones límite'.

Además, las leyes y la descripción deben *explicar* juntas la conducta de los miembros de la clase; el ejemplo de la clavija ya demostró que no cualquier deducción realizada por 'el fulano lógicamente omnisciente' sirve como explicación. En particular, lo anterior significa que, aunque no necesitamos que el comportamiento de un miembro de la clase natural pueda predecirse *realmente* en la práctica mediante una deducción factible, debe haber no obstante cierta categoría de hechos relacionados con esa conducta que cumpla con los criterios deseables y comunes para el caso de la explicación por medio de teorías científicas (generalidad, variedad significativa, significación y productividad de los resultados) y que sea deducible de las leyes y de la descripción junto con los juicios auxiliares correspondientes, y que esas deducciones satisfagan las restricciones de los 'intereses' que hemos descrito.

Lo expuesto constituye el ideal de las explicaciones en *física*, donde en ocasiones se lo lleva a cabo con un alto nivel de aproximación. (Un ejemplo sobre ese modelo explicativo, para el caso de la clase natural 'átomo de hidrógeno' estaría dado por las leyes de la mecánica cuántica y la explicación del átomo de hidrógeno como compuesto por un electrón y un protón en estado de enlace.)

Sin embargo, esa pareja formada por un conjunto de leyes y una descripción no resulta necesariamente *suficiente*. Lo que sigue *no* es un ejemplo sobre un modelo explicativo de la clase natural 'ser humano': las leyes de la física *más* la descripción de que 'un ser humano es un sistema físico que contiene partículas elementales'. ¿Pero por qué no? Después de todo, si el fisicalismo es verdadero, *ésta es* una descripción a partir de la cual, *dados los valores de los diversos parámetros* (es decir, *cuáles* son las partículas elementales y *cómo* están ordenadas, además de las condiciones elementales y las condiciones límite), el fulano lógicamente omnisciente podría deducir el comportamiento de cualquier miembro de esa clase.

El error reside en que se trata de un modelo explicativo de un género más amplio de 'sistemas físicos' (bajo una forma disfrazada) y no de la especie de los seres humanos. Para garantizar que un modelo de cualquier *género* (hasta los 'sistemas físicos' inclusive) no figure de un modo trivial como modelo explicativo de todas las especies, debemos exigir que sea verdadero el siguiente enunciado condicional subjuntivo: *cualquier cosa que pudiese satisfacer la descripción en cuestión sería un miembro de la clase natural a la que nos referimos*. Puesto que el enunciado indicativo de que 'cada sistema físico

es un ser humano' es falso, mucho más lo sería el condicional subjuntivo de que cualquier cosa que pudiera ser un sistema físico sería un ser humano, la pareja compuesta por las leyes de la física *más* el enunciado 'un ser humano es un sistema de partículas elementales' *no* cuenta como modelo explicativo del ser humano.

Por otro lado, con objeto de atender al hecho de que las leyes de la mecánica cuántica (o de la física según se la formule en *cualquier* época) son, sin duda, verdaderas sólo de manera aproximada y que, por lo tanto, puede *no* ser el caso que un sistema consistente en un protón y un electrón, *tal como los definimos ahora* sea un átomo de hidrógeno, se debe especificar que el principio del beneficio de la duda habrá de utilizarse para evaluar la verdad del enunciado condicional subjuntivo 'cualquier cosa que se componga de un electrón y un protón en estado de enlace sería un átomo de hidrógeno', o bien que ha de juzgarse la verdad 'desde el interior' del sitema conceptual al que pertenece la teoría en cuestión. El problema reside en que la oración condicional 'cualquier cosa que sea un sistema de partículas elementales constituiría un ser humano' no es verdadera, aun cuando aplicáramos el principio del beneficio de la duda a las expresiones 'sistema de partículas elementales' y 'ser humano', y tampoco resulta cierta si tomamos a la ciencia actual según 'su valor nominal', es decir, ignorando el hecho de que es sólo *aproximadamente* verdadera.

Cuando decimos que incluso los modelos explicativos ofrecidos por la física para clases naturales como la de 'átomo de hidrógeno' son correctos únicamente de manera aproximada, no debemos perder de vista el hecho de que, para la mayor parte de sus fines, el grado de aproximación puede ser, en realidad, muy alto. (Para el caso de los átomos de hidrógeno, es posible predecir los fenómenos con cifras de ocho decimales en la electrodinámica cuántica.)

Lo que sostengo, repito, es que existe a este respecto una diferencia fundamental entre la física y las ciencias sociales (o mejor, los *estudios* sociales). La primera proporciona, en efecto, modelos explicativos detallados para esas clases naturales. Pero no nos será posible obtener un modelo explicativo detallado que sea realista para la clase natural 'ser humano'. Y la dificultad no estriba *meramente* en su complejidad. Los meteorólogos cuentan con un modelo matemático del *clima* (o de muchos tipos de fenómenos climáticos), aunque evidentemente no están en condiciones de predecir el tiempo con toda exactitud en la práctica debido a que (a) los valores de

los parámetros requeridos no son conocidos con precisión en su totalidad, y (b) resultaría demasiado complejo llevar a cabo los cómputos en caso de que se los conociera. La razón de que podamos tener un modelo matemático bastante simple para la clase natural 'clima' radica en que, a pesar de que éste es un embrollo, se trata de un embrollo relativamente falto de estructuración. La causa de los problemas del teórico, a diferencia del ingeniero, no es la complejidad *per se*, sino la que está altamente estructurada, es decir, la complejidad de la estructura.

'En principio' y 'en la práctica'

Antes de discutir lo que podemos o no podemos hacer 'en principio', expliquemos lo que es razonable esperar que podamos hacer en la práctica. Aunque algunos creen que sería posible construir eventualmente un modelo explicativo detallado del ser humano, o por lo menos especificar las características de ese modelo que constituirían una descripción psicológica sobre la forma normal que debería presentar un ser humano (a lo que me he referido bajo el término de 'organización funcional' en la conferencia precedente), no se hallará un estudioso sensato de estas cuestiones que espere que lo logremos antes de trescientos años. 'Pero esto no es más que un simple hecho empírico.' Sí, es cierto que se trata de un hecho empírico, pero no tan 'simple'.

Los filósofos siempre han pretendido que ofrecen 'verdades necesarias', y les disgusta formular enunciados claramente 'empíricos'. Pero esto quizás sea una muy mala costumbre. Algunos hechos 'empíricos' pueden resultar *constitutivos* de nuestra naturaleza e instituciones presentes.

Supongamos, por lo tanto, que nuestra organización funcional se vuelve transparente a nuestros ojos. Supongamos que tenemos una teoría sobre esto y que la podemos *aplicar* efectivamente a una clase significativa de casos. ¿Qué nos pasaría? (Si contamos con esa teoría pero no la podemos usar para predecir y modificar el comportamiento, probablemente su significación sería muy escasa para nosotros.) ¿Resultaría posible *amar* a alguien, si fuéramos en verdad capaces de efectuar este tipo de *cálculos*: 'si digo *X* la probabilidad de que ella reaccione de la manera *Y* es del quince por ciento'? ¿Sería posible tener amigos y enemigos? ¿O incluso pensarse a *uno mismo* como *persona*? Ignoro la respuesta a estas preguntas, pero me

parece evidente que el desarrollo de este tipo de conocimiento sobre nosotros mismos y sobre los demás modificaría nuestras naturalezas en formas que no podemos predecir en modo alguno. Todas las instituciones con las que ahora contamos (arte, política, religión, e incluso la ciencia) quedarían tan alteradas que no las reconoceríamos.

Pero un hecho, por 'empírico' que sea, que fundamenta a la totalidad de nuestra historia y que influye en el carácter de cada una de nuestras instituciones no es 'simple'. El que resultemos parcialmente oscuros para nosotros mismos, en el sentido de que carecemos de la capacidad para entender a los demás como entendemos a los átomos de hidrógeno, es uno de esos hechos, a saber, un hecho *constitutivo*. Si el mismo ya no prevalece después de, digamos, trescientos años, me da igual: no estoy escribiendo filosofía para la próxima edad histórica ni para nuestros descendientes post-humanos, sino para los seres humanos de la actualidad.

Qué significa 'en principio' y qué debería significar

Por lo general, los filósofos sólo distinguen dos o tres nociones de posibilidad: lo que es posible en el sentido 'lógico'; lo que es posible físicamente en el sentido de 'que no contradice ninguna ley natural'; y lo que es posible técnicamente dada la tecnología existente. '¿Es posible 'en principio' hacer *X*?' significa por lo regular: '¿se puede hacer *X* desde el punto de vista físico (o quizás, lógico)?' Pero esto no es lo que debería significar en las argumentaciones de temas como el que ahora estamos exponiendo. Existe una importante noción 'intermedia' de la posibilidad (entre la física y la técnica), a la cual se ha pasado por alto.

Para definir esta acepción, consideremos el problema que se discute con frecuencia acerca de si la conducta humana puede *predecirse* 'en principio' (dentro de los límites fijados por la indeterminación de la mecánica cuántica). La pregunta pertinente no consiste en si el Fulano Lógicamente Omnisciente puede lograrlo (es decir, si tiene posibilidades lógicas de hacerlo), ni siquiera si resulta posible desde el punto de vista físico para un sistema computarizado, sino más bien si es físicamente posible efectuar esto *con la rapidez suficiente para que se lo considere una predicción*, si cabe realizar ese cálculo 'en el tiempo real', como dice la gente que trabaja en computación. Y aunque de 'la hipótesis del fisicalismo' se concluya que debe ser posible para el Fulano Lógicamente Omnisciente predecir nuestra conducta

a partir de una teoría ideal (dentro de los límites establecidos por la indeterminación de los procesos elementales) e incluso que debe ser físicamente posible (aunque no lo creo), no hay razón para pensar sobre esa base, o cualquier otra, que se lo pueda efectuar *en el tiempo real*. Quizás ocurra, por ejemplo, que la deducción más breve de lo que voy a hacer dentro de cinco minutos (o las probabilidades de sus diversas opciones) *requiera más de mil años para ser computada* en un sistema de computación óptimo que utilice una teoría ideal acerca de mí y un programa óptimo para verificar los teoremas. En tal caso ésta representaría una acepción de 'en principio' en la cual mi conducta no sería predecible *en principio* de esta manera, aun cuando el Fulano Lógicamente Omnisciente hiciera la predicción (indeterminista). La ciencia de la computación nos ha dado un nuevo concepto de 'en principio' con su noción de 'computabilidad en el tiempo real', la cual resulta mucho más pertinente para las discusiones sobre predictabilidad que las nociones corrientes.

Lo expuesto con respecto a la predicción también se aplica a la explicación. La organización funcional humana evolucionó a lo largo de uno o dos millones de años. No hay razón para que la humanidad posea una descripción que quepa en un solo libro, ni siquiera en la Biblioteca Bodleiana. Hablar de explicación, no la de algunos hechos muy generales acerca de la psicología humana, sino de fenómenos específicos como la causa de que algunas explicaciones nos parezcan naturales y otras no, o de que algunos puntos de vista no nos resulten 'razonables', o deducir siquiera una caracterización de esas cuestiones a partir de una descripción hipotética de nuestra organización funcional, requeriría tanto 'tiempo real' que la razón humana habría quizás dejado de existir mucho tiempo antes de que la deducción estuviera terminada.

El problema radica en que nuestra mejor teoría actual sobre el tema sugiere que ciertas habilidades humanas —el lenguaje hablado es el ejemplo paradigmático— tal vez no sean explicables teóricamente de manera aislada. Sin duda, es casi imposible elaborar un 'modelo' del hablante de un lenguaje sin efectuar un modelo de toda la organización funcional humana. Pero esta última bien podría resultar *ininteligible* para los seres humanos, una vez que esté formulada en detalle. La moraleja no consiste en que no podamos estudiarnos a nosotros mismos, sino que (en la práctica y, según mi argumentación, muy posiblemente también *en principio*) no podemos estudiarnos de la manera en que estudiamos los átomos de hidrógeno.

SEXTA CONFERENCIA

Hoy quiero exponer algunas cuestiones generales que conciernen a los métodos y el carácter de las ciencias sociales. Los filósofos empiristas de la ciencia, desde Mill hasta los empiristas lógicos, han tendido a considerar la física como la idea platónica de la ciencia, y a esperar el día feliz en que las ciencias sociales, por la gracia del 'método científico' se parezcan a la física. El espíritu quedó fijado con la afirmación de Mill de que 'el estado de atraso de las ciencias morales sólo puede remediarse aplicándoles los métodos de la ciencia física, debidamente extendidos y generalizados'.[1] Y esta propuesta se vio reafirmada, con algunos matices modernos, en las aseveraciones de Zilsel, según las cuales 'sólo se descubrirán las leyes de la historia cuando se recoja y compare el material con minuciosidad filológica y no mediante discusiones metodológicas generales como las nuestras',[2] y 'la física es la más madura de todas las ciencias empíricas en cuanto a su método. En física, el concepto de ley ha sido usado durante trescientos años. Debe asumirse, por lo tanto, que la mayor parte de las dificultades de su aplicación a otros campos tiene su contraparte en esa ciencia y puede clarificarse fácilmente con la ayuda de conceptos físicos';[3] y también en la afirmación de Nagel de que 'no parece haber un buen motivo para pretender que el patrón general de las explicaciones en la investigación histórica, o la estructura lógica de las herramientas conceptuales empleadas en

[1] En *A System of Logic*. Las palabras citadas son las que abren el libro VI, octava edición, Nueva York, 1881.

[2] Edgar Zilsel, 'Physics and the Problem of Historico-Sociological Laws', en H. Feigl y M. Brodbeck (eds.), *Readings in the Philosophy of Science*, Nueva York, Appleton-Century-Crofts, 1953, pp. 714–722.

[3] *Ibid.*, p. 714.

ella, difieran de las que se encuentran en las ciencias de la generalización y natural'.[4] El énfasis en la física se revela por el hecho de que Nagel, inmediatamente después de la afirmación citada, continúa diciendo que las 'premisas explicativas en la historia' incluyen las leyes 'así como también muchos juicios singulares formulados de manera explícita (aunque, por lo general, incompletos) sobre las condiciones iniciales'. 'Ley' y 'condiciones iniciales' son términos que pertenecen, en última instancia, a la jerga de los físicos y no es una forma de hablar de los historiadores.

Como cabía esperar, la pretensión de algunos de los más grandes sociólogos del siglo XIX de que *Verstehen —o la capacidad para imaginar cómo habría sido*, por ejemplo, un caballero del ejército de Ricardo Corazón de León o un seguidor de Savonarola— es, en tanto opuesto a nivel metodológico a la mera significación psicológica en las ciencias sociales, como un anatema para Nagel *et al.*

De acuerdo con estos filósofos empiristas, *Verstehen* (la comprensión empática) puede ser una fuente de hipótesis en las ciencias sociales, pero no un método de confirmación. Las hipótesis deben ser comprobadas mediante la formulación de predicciones, la comparación con los datos, etc. En el caso de la historia, deben incluir o presuponer *leyes generales* que se comprueban a su vez con la deducción de predicciones, etc. Las dos ideas principales son:

(I) Los métodos de la física (como los describen Mill, Nagel, etc.) son los métodos —los únicos métodos— de *todas* las ciencias.

(II) Conocimiento = ciencia = el 'método científico', es decir, cualquier cosa que pueda conocerse debe ser conocida por estos métodos. El hipotético deductivo y sus parientes inductivos más simples (junto con la observación, por supuesto) son todos los métodos que existen para obtener un conocimiento no demostrativo.

Me referiré primero al punto (II), el cual, por cierto, es ambiguo: podría significar que, como *una hipótesis de psicología cognoscitiva*, todo lo que conocemos se aprende a través de un proceso que concuerda con los paradigmas de la lógica inductiva, posiblemente un proceso inconsciente (al que incluso quizás no fuéramos capaces de

[4] Ernest Nagel, "The Logic of Historical Analysis", en H. Feigl y M. Brodbeck (eds.), *Readings in the Philosophy of Science*, Nueva York, Appleton-Century-Crofts, 1953, pp. 688–700.

poner por escrito en su totalidad). O bien, podría significar que todo lo que conocemos *puede comprobarse* siguiendo los paradigmas de la lógica inductiva.

Ernest Nagel vacila curiosamente en esta precisa cuestión. Escribe:

> No hay duda de que el problema básico en este campo de investigación consiste en que no tenemos en la actualidad un modelo para sopesar los datos de una hipótesis determinada arbitrariamente que se acepte a nivel general, que esté formulado en términos explícitos y que lo abarque todo, de tal forma que pueda compararse el valor lógico de las conclusiones alternas con los datos disponibles para cada una. Incluso en problemas de gran importancia práctica, los juicios deben formularse con base en consideraciones que sólo se comprenden de una manera vaga y, como carecemos de un canon lógico convencional para calcular en qué medida los datos justifican una conclusión, cuando los juicios se contradicen entre sí, ello a menudo parece ser consecuencia de un procedimiento esencialmente arbitrario. Esta circunstancia afecta tanto a la posición que tienen las conclusiones del historiador como a los descubrimientos de otros estudiosos. Por fortuna, aunque la gama de desacuerdos posibles con respecto a la importancia que poseen los datos para un juicio dado es teóricamente ilimitada, existe un punto de coincidencia sustancial entre los hombres con experiencia en los temas pertinentes, acerca de las posibilidades relativas que deben asignarse a muchas de las hipótesis. Este punto de acuerdo indica que, a pesar de la falta de una lógica expresada en términos explícitos, gran parte de los hábitos tácitos de pensamiento comprenden principios de inferencia que pueden justificarse fácticamente. En consecuencia, aunque a menudo existen razones legítimas para dudar de la validez de algunas asignaciones causales específicas que se efectúan en historia, no parece haber motivos obligados para convertir tal duda en un escepticismo indiscriminado.[5]

Por un lado, esta cita puede interpretarse como una hipótesis sobre nuestros 'hábitos tácitos de pensamiento' (y sobre esos 'principios de inferencia que pueden justificarse fácticamente' y que no sabemos qué son). ¿Quién negará que los 'hábitos tácitos del pensamiento' de un buen historiador comprenden *unos* u otros 'principios de inferencia que pueden justificarse fácticamente'? Tal vez 'basarse hasta cierto grado en el *Verstehen*' sea un 'principio de inferencia jus-

[5] *Ibid.*, p. 700.

tificable fácticamente'. Por otro lado, los calificativos 'en la actualidad' y 'formulado en términos explícitos' (en la afirmación de que 'no tenemos actualmente un modelo para sopesar los datos [...] que se acepte a nivel general, *que esté formulado en términos explícitos* y que lo abarque todo') sugieren que toda ciencia —incluyendo la historia, que es el tema del artículo de Nagel— *puede* ser construida siguiendo un 'canon lógico convencional', el cual representa en sí mismo una reelaboración de la explicación convencional de la metodología usada en 'las ciencias de la generalización y natural', es decir, de la *física*. Tomaré esta segunda interpretación de las palabras de Nagel, dado que sólo en este sentido habla de metodología, sin especular acerca de nuestros 'hábitos de pensamiento' inconscientes.

Pero bajo *esta* interpretación, el enunciado (II) resulta absurdo. Y no es preciso que nos compliquemos en la gran polémica acerca de las 'leyes generales en historia' para darnos cuenta. Por el momento, limitémonos al tema sobre el que tanto hemos hablado en estas conferencias, el de la *traducción*.

En temporada reciente, estuve estudiando hebreo. Así que supondré que todos sabemos que la palabra *shemen* significa aceite (Quine negaría que esto forma parte del *conocimiento*, pero en este punto se encuentra prácticamente solo). ¿Puede este ejemplo de conocimiento ser comprobado 'científicamente' en el sentido de Mill o Nagel?

Dado que una manera de comprobar esa afirmación consiste en *preguntar a las personas bilingües*, y hay mucha gente que habla hebreo y español, parecería no existir problema. Preguntemos a cien bilingües (o a mil o al número que sea). *¿Pero cómo lo sabe esta gente?* (La pregunta: *¿lo sabrán realmente?* es un tipo especial de escepticismo.)

Si no quiero recurrir a las personas bilingües, sólo tengo que aprender hebreo (de manera ideal, en Israel por el método directo, es decir, sin depender de las explicaciones en español). Pero ¿cómo compruebo lo que pienso que estoy aprendiendo?

Bien, si voy a una gasolinera y digo *bedok et hashemen* y el despachador me da un puñetazo en la nariz (y lo mismo me pasa en otras gasolineras) mi fe en que la traducción de *bedok et hashemen* es 'revise el aceite' se tambalea. Pero observemos lo que está pasando. Estoy presuponiendo: (1) que el empleado quiere vender gasolina

y aceite; (2) que no es *obligatorio* en Israel decir *bevakasha* (por favor) cuando uno solicita algo; (3) que si alguien desea vender aceite y el cliente dice 'revise el aceite' en el lenguaje del vendedor (y ninguna regla obligatoria de educación ha sido violada), éste revisará el aceite (o tal vez contestará: 'no tengo aceite' en su idioma o, con menos frecuencia, 'estoy muy ocupado', pero *no* le dará un puñetazo en la nariz al cliente); (4) que una persona que llega a una gasolinera será tratada como cliente. Cada vez que revise mis 'hipótesis analíticas' (es decir, mis habilidades como traductor) en un contexto distinto, se añadirá una *nueva* lista de hipótesis psico–sociológicas del tipo (1)-(4), importadas del 'conocimiento general del entorno'. Resulta obvio que la lista completa de las cosas que uso y en las que creo no puede *redactarse* por anticipado.

Aunque parezca extraño (porque todo esto lleva agua a su molino) Quine simplifica excesivamente lo anterior con su noción de 'significado motivador'. Desde su punto de vista, sólo importan los contextos en que *yo* enuncio una oración y el hablante hebreo contesta *sí* (*ken*) o *no* (*lo*). Pero en realidad, aun si hubiera que adaptar a la perfección una hipótesis analítica a *esas* situaciones, es posible, al menos desde la perspectiva lógica, que la misma se derrumbe por completo cuando se la aplique 'en la calle' (o en una gasolinera), es decir, al tratar de mantener una conversación o 'comprender' alguna charla escuchada por casualidad. Y ésta es, sin duda, la prueba que nos preocupa.

Moraleja: probar la 'hipótesis de que *shemen* significa *aceite*' implica la comprobación *conjunta* de una 'hipótesis analítica' *que no puedo formular* apropiadamente (como lo demuestra el fracaso de la traducción mecánica) y unas premisas psicológicas que, como resulta evidente, no puedo enumerar por anticipado.

En conclusión, el conocimiento de un hecho tan *simple* como que '*shemen* significa aceite' no puede ser justificado o confirmado siguiendo los paradigmas de la lógica inductiva. Tal vez las inferencias que estamos realizando pueden ser convertidas en un paradigma por un marciano (o por una supermente laplaciana), pero los *seres humanos* no las podemos volver explícitas (al menos no en la práctica, y quizás ni siquiera 'en principio'). Tenemos que basarnos en supuestos como: 'Bueno, he aprendido hebreo (o cualquier lenguaje) y estoy seguro de que lo entiendo bien porque puedo hablar con la gente y comprender las respuestas que me dan.'

Pero el saber que uno *está* 'hablándole a la gente y entendiendo sus respuestas' constituye el típico conocimiento *práctico*. Las inferencias no son, en un sentido *serio* y no trivial, inferencias *científicas* (excepto, *tal vez* 'a nivel inconsciente'), y los términos clave —'significa', 'entiende'— no pueden, como ya he argumentado, precisarse científicamente.

Me parece que hay una cierta versión del cientificismo en las ciencias sociales que se derrumba justo en este punto. La idea de que aquello que *sabemos* es coextensivo a lo que estamos en posibilidades de confirmar 'públicamente' siguiendo paradigmas bien establecidos de comprobación científica ni siquiera sirve para algunos de los hechos *más simples* que conocemos, como el significado de las palabras en una lengua extranjera.

(Tal vez éste es el verdadero motivo de que Quine crea en la indeterminación de la traducción: desde su perspectiva, dado que este conocimiento no pertenece a la 'ciencia' en el sentido del empirismo lógico, no es conocimiento.)

Digresión sobre el criterio del consenso

Parte de la doctrina propugnada por los empiristas lógicos es que *conocimiento = conocimiento público*. Si ello implica que lo conocido es algo que puede ser aceptado por todas las personas competentes (¿o casi todas? ¡*todas* es demasiado escrupuloso!) que estudian el tema (donde la competencia se mide de acuerdo con ciertas aptitudes, dependiendo del campo de estudio), entonces el ejemplo de '*shemen* significa aceite' constituye un caso de conocimiento *público* que no emplea los paradigmas de la lógica inductiva de manera explícita (y no puede hacerlo, desde el punto de vista práctico). Además, con este criterio (el consenso de los competentes), la filosofía escolástica de la Edad Media también habría sido 'pública'.

Si, por otra parte, 'conocimiento = conocimiento público' significa que toda la gente *idealmente racional* estará de acuerdo, entonces resulta demasiado vago. 'La gente que sigue los paradigmas de la lógica inductiva de un modo inteligente, ¿estaría de acuerdo en su totalidad?' Ya he señalado que ello *no puede* aplicarse (explícitamente) a esta clase de tema. En cualquier caso, *no* es una verdad necesaria que toda la gente idealmente racional tenga que coincidir con respecto al conocimiento. *Mi conocimiento* es que desayuné

cereales esta mañana. Pero supongamos que soy un mentiroso empedernido. Quizás entonces las personas idealmente racionales no acepten lo que digo. Tal vez no puedan saber si comí cereal o no. Sin embargo, sigue siendo mi *conocimiento*.

Conocimiento práctico

¿De dónde *proviene* el hecho de que tengamos un conocimiento (incluso un conocimiento *público*, como acabamos de ver) que no puede ser 'transformado en ciencia' o 'verificado', mediante una verificación que concuerde públicamente con el criterio de 'metodología científica'? Resulta evidente que una de sus procedencias es la que sigue: que podemos adquirir *habilidades* que son demasiado complejas como para que una teoría las describa. Puedo aprender a traducir de un lenguaje a otro, pero no me es posible describir la habilidad que he adquirido mediante una teoría (explícita). Tal vez mi *cerebro* posee alguna teoría. Dado que el lenguaje tiene una estructura discreta, mi cerebro debe emplear algo semejante a los algoritmos, aunque no creo que éstos sean lo *único* que interviene, aun cuando se trate de una traducción. Sin embargo, suponiendo que mi cerebro posee una 'hipótesis analítica' completa 'escrita' en un hipotético 'lenguaje cerebral', *yo* no la tengo, y tampoco los científicos del presente. El hecho principal es que las *habilidades* no siempre dependen de *teorías* (como lo demuestra el ejemplo de *caminar*). Y el conocimiento, incluso el verbalizado, puede estar encarnado en una *habilidad* y no en una teoría.

Las explicaciones psicológicas comunes representan otro caso del mismo fenómeno. Todos los que participan en determinada situación quizás se dan perfecta cuenta de que Jones está celoso de la reputación de Smith. Pero no es posible ofrecer *nada que se parezca* a una 'prueba científica' de ello. No se trata, por ejemplo, de que 'Jones dijo bla-bla-bla, y generalmente la gente que dice bla-bla-bla tiene celos'. Aun si esto último fuera verdad, se pueden concebir sin dificultades *un número indefinido* de situaciones en las cuales alguien puede decir bla-bla-bla y *no* estar celoso. Por lo tanto, se trata más bien de que: 'es probable que la gente que dice bla-bla-bla esté celosa *a menos que prevalezcan circunstancias especiales* y no hay circunstancias especiales en este ejemplo' (una inferencia *ceteris paribus*, como el silogismo práctico). El caso se parece al problema de

traducción en que no se puede 'verificar' aisladamente *Jones está celoso*: habría que verificar una *amplísima* 'teoría psicológica' que cubriera todas las 'circunstancias especiales'. Y esto, por supuesto, se encuentra *implícito* en nuestro conocimiento de la gente y en nuestra *habilidad* para utilizar descripciones psicológicas; no es algo que podamos enunciar en términos explícitos. Se trata otra vez de una característica del conocimiento 'científico' (por lo menos, si tomamos a la *física* como paradigma) el hecho de *emplear instrumentos de medición que nos resulten comprensibles*. Nuestra teoría se aplica a los instrumentos de medición que poseemos, y a sus interacciones con lo que éstos suelen medir, y no sólo a los objetos que medimos. Es característica del conocimiento *práctico* el que a menudo debamos utilizarnos a nosotros mismos (o a otra persona) como instrumentos de medición... y que *carezcamos* de una teoría explícita para *éstas* interacciones.

Física y lógica

Sin embargo, incluso en la física, la situación no es tan simple. En realidad, lo que la teoría describe es (normalmente) un 'sistema cerrado' idealizado. La teoría de este 'sistema cerrado' puede ser tan precisa como se quiera. Y es *dentro* de esta idealización que se obtienen los conocidos ejemplos sobre el 'método científico'. Pero la *aplicación* de la física depende del hecho de que podamos producir en el laboratorio o encontrar en el mundo sistemas *abiertos* que se aproximen al sistema idealizado en grado suficiente como para que la teoría arroje predicciones muy precisas. La decisión de que, en un caso determinado, las condiciones se han aproximado lo bastante —tanto que incluso *vale la pena* aplicar el modelo idealizado a *este* caso— depende comúnmente del conocimiento práctico no formalizado. La moraleja es que el supuesto 'método científico' ('MC') sólo constituye una formalización de *algunos aspectos* de la metodología científica. La física misma no podría avanzar si utilizara *únicamente* el 'MC'.

La distinción entre el conocimiento obtenido por el 'MC' y el conocimiento práctico no es igual entonces a la división entre *ciencia* y otros tipos de *conocimiento*. La ciencia *presupone* cierto 'conocimiento no científico'. Por ejemplo, he afirmado que 'se refiere' y 'verdadero' (es decir, se refiere en L y verdadero en L para L variable) no pueden 'precisarse científicamente'. Sin embargo, la *verdad*

constituye un término fundamental en *lógica*, que es una *ciencia precisa*.

Por otra parte, un historiador o un sociólogo tiene que basarse en *traducciones* de documentos extranjeros, y por lo tanto en un conocimiento *no* comprobado (o no comprobable, al menos en la práctica) por el 'MC'. Pero aún más: si un historiador revisa documentos, examina acciones públicas, lee diarios y cartas, ¿cómo va a decidir entonces si 'Smith estaba hambriento de poder'? *No* mediante la aplicación a esos datos de las 'leyes generales de la historia, la sociología y la psicología', como lo exhortan a hacer los metodólogos positivistas. Más bien debe absorber todo este material y luego *confiar en su sabiduría humana* para decidir que esto demuestra hambre de poder 'más allá de cualquier duda razonable' (como se dice en los tribunales). En efecto, se emplea *a sí mismo* como 'instrumento de medición', lo cual representa bastante más de lo propuesto por Weber, Dilthey, etc. (me explayaré sobre el *Verstehen* más adelante). Algunas veces su decisión resulta controvertida (aun cuando sea correcta), pero otras no (¿puede alguien dudar de que *Nerón* era un ególatra?).

Desde luego, el punto aquí señalado es la similitud entre el conocimiento de la física y el de las ciencias sociales, es decir, que ambos dependen de un conocimiento práctico no formalizado, lo que Polanyi denomina 'conocimiento implícito'. Pero tal similitud se funda y presupone una *diferencia*. En mi conferencia anterior argumenté que, aunque en algunas ocasiones logramos construir buenos modelos explicativos para ciertas clases naturales de la física, resulta inútil buscar un *modelo explicativo* para la clase natural 'ser humano' (aunque podríamos, por supuesto, descubrir regularidades empíricas y tal vez incluso *leyes* sobre los seres humanos). Esto explica la *imposibilidad de formalizar* al conocimiento práctico.

Algunas cuestiones psicológicas

Ahora quiero hacer una digresión sobre algunas cuestiones psicológicas.

Consideremos el caso de pensar en cómo se escala una montaña 'escalándola en la imaginación'. Esto sugiere claramente que mucha de la información con que contamos puede no hallarse *almacenada* en forma de proposiciones. Podemos *enunciar* las consideraciones

pertinentes con palabras, pero ello no significa que estuvieran formuladas en un 'código del cerebro'. (De hecho, hay una buena cantidad de experimentos que indican firmemente que el cerebro almacena *imágenes*.) No estoy afirmando que este órgano sea 'analógico', en oposición a una máquina 'digital'; más bien sugiero que no es *ni lo uno ni lo otro*. Para cualquier teoría verosímil, el cerebro presenta elementos digitales (se producen cosas parecidas a las que ocurren en una computadora); pero bien puede procesar una cantidad considerable de información que no es digital. En particular, la *traducción* no necesita ser 'mecánica' en todos sus aspectos. Puede suceder que a veces asociemos imágenes con las situaciones en que nos encontramos, a fin de descubrir el 'marco' que debemos usar en la traducción. En tal caso, el examen de nuestra habilidad para traducir no consistiría en probar una *teoría*, ni siquiera una 'inconsciente'; y se puede decir lo mismo con respecto a nuestra capacidad de explicación psicológica e incluso de explicación de los objetos materiales de nuestro lenguaje ordinario.

En segundo lugar, algunas de nuestras habilidades para *imaginar* las reacciones probables de la gente, bien pueden ser *innatas*. De hecho, nuestra disposición para *creer en lo que nos dicen otras personas* difícilmente es producto de una inferencia; de otro modo, como lo advierte John McDowell, no tendríamos ninguna justificación para decir que *conocemos* lo que hemos aprendido de los demás. (Si se trata de una inferencia, no resulta demasiado buena.) Pero la disposición para creer lo que se nos dice constituye a la vez un caso especial de *Verstehen* y la base de todo conocimiento. Y no sólo del conocimiento social; la física puede fundarse en el *Verstehen* tanto como las ciencias sociales.

Verstehen

Ahora bien, los críticos del *Verstehen* dicen que 'ponerse en el lugar de otro' puede ser una fuente útil de *hipótesis*, pero éstas todavía debe ser *confirmadas* (usando el 'MC'). Aquí he criticado parte de esa afirmación: el *control* de tales hipótesis no se efectúa usando el 'MC' de manera explícita y paradigmática, y tal vez ni siquiera 'inconsciente' (aunque ésta es más una cuestión de psicología cognoscitiva que metodológica). Expondré ahora la otra parte: la opinión de que el *Verstehen* pertenece al 'contexto del descubrimiento' y no al 'contexto de la justificación'. Supondré que lo 'conocido' por em-

patía necesita ser controlado y, por lo tanto, no es *conocimiento*. De ello no se concluye que *sólo* se trata de una 'hipótesis'. En Filosofía de la Ciencia se acepta la doctrina de que la comprobación inductiva de las teorías presupone un sopesamiento *a priori* (en el sentido de que es precedente) de las teorías, el cual es un procedimiento *anterior* al control. En las teorías probabilistas de la confirmación (las que están basadas en el teorema de Bayes) esto podría ser una 'distribución de probabilidad previa'; en las teorías por eliminación se trataría de un 'ordenamiento según la simplicidad'. (Popper piensa que es posible emplear sencillamente una 'medida de contenido', pero esto no sirve.) Aquí aludiré a la 'probabilidad previa', aunque yo mismo pienso que se trata sólo de un término cualitativo, no numérico.

También se acepta que nadie ha sido capaz de *formalizar* los juicios inductivos de los científicos que *en la práctica* constituyen la 'métrica de la probabilidad previa'.

Ahora bien, me interesa sugerir que la empatía proporciona quizás algo inferior al 'Conocimiento' (con C mayúscula), pero superior a la mera posibilidad física o lógica. Ofrece *verosimilitud*, la cual constituye la fuente de la probabilidad previa para muchos de los juicios sobre las personas. Reviviendo la terminología platónica, quizás no suministre 'conocimiento', pero sí 'opiniones correctas'; y estoy afirmando que el conocimiento *depende* en alto grado de la opinión correcta.

Me parece que esto es todo lo que un defensor inteligente del *Verstehen* en las ciencias sociales tiene que decir contra Nagel *et al.*. El *Verstehen* es una fuente de la *probabilidad previa*; de hecho, la 'hipótesis' a la que asignamos una probabilidad previa muy elevada con base en el *Verstehen* (empatía) debe ser 'controlada', pero *incluso este control es finalmente intuitivo*. La imagen alternativa de la ciencia social —la que se funda en las 'leyes generales de la historia, la sociología y la psicología', confirmadas por el 'MC' y aplicadas a los datos, como en la física— resulta falsa para la ciencia social del presente y utópica como visión de su futuro.

El propósito de Nagel *et al.* consistía en eliminar lo que consideraban el 'oscurantismo' y la 'metafísica' de las ciencias sociales. Aunque tales fines sean valiosos, perseguirlos mediante la tergiversación de los sucesos reales equivale a promulgar una 'ideología', en el sentido peyorativo del término, y no a propiciar la claridad metodológica.

¿Recomendaciones?

Muchos científicos sociales, y muchos que no lo son, pueden molestarse con lo que he dicho. 'Muy bien', tal vez deseen replicar, 'nos ha convencido de que las ciencias sociales no pueden aspirar de manera realista a la misma clase de modelo matemático detallado para su materia de estudio que los físicos tratan de construir para la suya, y que seguirán siendo indispensables diversos tipos de conocimiento intuitivo y no formalizado. ¿Pero qué recomienda usted en lugar de esa imagen de las ciencias sociales que intentan parecerse más a la física?' Esta pregunta surge del hecho real de que la sociología, en particular, descubrió a fines del siglo XIX que le resultaba cada vez más difícil alcanzar algún tipo de objetividad, y que incluso los historiadores y los politólogos de hoy deben soportar que se les objete que 'ésa es sólo *su* interpretación'. ¿Qué haremos sin el criterio de objetividad suministrado por la emulación de la física?

Es cierto que la metodología empirista buscaba sacarnos de las oscuridades del desacuerdo y la controversia. Pero en ocasiones, cuando se ha tomado el camino equivocado, la mejor forma de salir de una selva sombría es volver a entrar en ella. No sé qué maravillas lograrán en el futuro las ciencias sociales (aunque no veo razón para dudar de que la inteligencia, la imaginación y la sensibilidad producirán en el porvenir obras maestras en este campo, como lo han hecho en otros), pero sé que 'cientifizarlas' (un término bárbaro que inventé para expresar una idea bárbara) constituye una confusión y una fuente de confusiones.

Si a fin de cuentas *no* tengo recomendaciones específicas que hacer a los científicos sociales —o más bien a los estudiosos de los fenómenos sociales y humanos—, puedo de todas formas anticipar algunos malentendidos. *No* estoy diciendo que las matemáticas no tengan un lugar en las ciencias sociales. La economía matemática no nos presenta un modelo matemático de la persona total; pero, aun así, su labor es valiosa. No estoy afirmando que las teorías verificables carezcan de un lugar en ellas, o la estadística o el 'MC'. Mi concepción es más bien pluralista. Todos esos campos poseen un lugar y debemos usarlos cuando y donde resulte posible. Pero a esas técnicas no les faltan defensores. El propósito de esta conferencia fue restablecer el equilibrio subrayando los derechos de la vasta reserva de conocimiento humano no formalizado y no formalizable,

de la cual dependemos y con la cual vivimos, respiramos y existimos cada día de nuestras vidas.

Si se admitiera el punto de vista por el que estoy abogando, sin duda se aceptarían ampliamente tanto la diversidad como la continuidad de los estudios sociales y humanos; por ejemplo, la continuidad y la diversidad que hay entre los estudios empíricos en sociología y en historia (ya no concebidas como consistentes en 'leyes' y 'condiciones iniciales'); o la continuidad y la diversidad existentes entre, por un lado, la historia y la sociología y, por el otro, el ensayo político y el cultural. Las ciencias sociales tal vez se volvieran, en parte, más 'literarias'. ¡Incluso algunos libros podrían estar bien escritos! Habría menos preocupación por encontrar una regla metodológica capaz de eliminar las controversias y lograr el consenso. En realidad, para determinados temas, el único consenso que debería obtenerse es el de los sabios.

¿Tenemos que *lamentar* el hecho de que las ciencias sociales no puedan aspirar de manera realista a parecerse a la ciencia física? Preguntar esto equivale a inquirir si debemos lamentar el hecho de que no podamos conocernos a nosotros mismos ni a los demás de la misma forma en que el físico entiende al oscilador armónico (ni de la manera en que Dios podría vernos, la cual es seguramente muy distinta). ¿Pero es un destino tan terrible el de hallarnos condenados a no tener sobre nosotros mismos ni la perspectiva de la computadora ni la perspectiva de Dios? Somos hombres y mujeres, y espero que contemos con la suerte suficiente para seguir siendo hombres y mujeres. Tratemos de preservar nuestra humanidad asumiendo, entre otras cosas, un punto de vista humano con respecto a nosotros mismos y al conocimiento de nuestro ser.

SEGUNDA PARTE

LITERATURA, CIENCIA Y REFLEXIÓN

LITERATURA, CIENCIA Y REFLEXIÓN*

Aunque mi intención es analizar la relación de las formas de pensamiento literario y científico con lo que puede ser apropiadamente denominado la comprensión humana, por razones que se esclarecerán en breve, comenzaré por hablar un poco de esa rama de mi propia materia que es la filosofía moral. Lo que diré sobre esta última se halla muy influido por una obra reciente, y desafortunadamente aún inédita, de Judy Baker y Paul Grice. Grice y Baker asumen que el problema es una simple pregunta de cuatro palabras: '¿cómo hay que vivir?' Me gusta este punto de partida, no sólo porque parece fresco y poco trillado en comparación con las formas académicas usuales de empezar una investigación sobre temas éticos, sino porque además produce el efecto de convertir inmediatamente a la ética en una rama del conocimiento práctico, y creo que ésta es una idea muy importante.

Pero, ¿es posible discutir en términos racionales la pregunta 'cómo hay que vivir'? Por supuesto, la persona que posee una moral, por lo menos en el sentido de haber llegado a un 'estilo' o conjunto de hábitos de vida, e incluso quien ha reflexionado sobre su estilo de vida y es consciente de tener lo que llamamos una personalidad, y que reconoce o piensa que debe aceptar ciertos tipos de crítica a su carácter, por ejemplo, que ha elegido de manera irracional o estúpida los medios para lograr ciertos fines, puede rechazar una crítica más fundamental de su propia moral o personalidad diciendo simplemente: 'bueno, así es como quiero vivir' o 'esta forma de vivir me satisface'. Si la mayoría de la gente bloqueara de este modo todas las críticas, excepto las objeciones más superficiales contra sus

* Reproducido, con la amable autorización del editor, de *New Literary*, vol. VII, 1975–1976.

medios y fines, no habría surgido, ni podría surgir, una institución como la moral.

Me parece que un hecho psicológico muy importante, aunque desafortunadamente no es de una clase a la que los psicólogos se refieran con frecuencia, consiste en que la gente reflexiona sobre su propia personalidad, y también que por lo general trata de justificarla ante los demás, por lo menos cuando recibe una crítica. A la mayoría de nosotros nos interesa justificar siquiera algunas características de nuestro estilo de vida, en el sentido de presentar una defensa de ellas que interese a los demás como justificación. Claro que hay límites. Como también apuntan Baker y Grice, no sólo queremos tener una moral que resulte muy atractiva, sino que nos permita al mismo tiempo un gran 'espacio discrecional'. Una moral que nos dictara un deber en cada circunstancia concebible sería invivible.

No obstante, lo cierto es que cuando vemos que el razonamiento moral no tiene lugar en el vacío cartesiano, sino entre gente que trata de responder a las críticas contra su personalidad, y entre gente que intenta justificar sus formas de vida ante otros que buscan criticar la forma de vida de algunos más, etc., aduciendo razones que tienen cierto tipo de atractivo general, nos damos cuenta de que la cuestión de la objetividad en la ética surge de una manera enteramente nueva o que aparece bajo una luz por completo distinta. La cuestión de si hay una o varias morales que sean mejores en términos objetivos y que coincidan, prometedoramente, en una buena cantidad de principios o de casos, se reduce sencillamente al problema de si, dadas las aspiraciones que nacen de forma automática cuando nos dedicamos a elaborar una justificación de atractivo *general* para los principios de vida, tales aspiraciones seleccionarán la mejor moral o un conjunto de morales capaz de lograr puntos de acuerdo considerables para una serie de cuestiones importantes.

En la terminología utilizada por John Rawls, el problema consiste en si existe un conjunto de morales o una única moral que pueda pasar la prueba del 'equilibrio reflexivo general'. De hecho, se podría considerar que Grice y Baker están tratando de obtener una descripción detallada del proceso mediante el cual 'se llega al equilibrio reflexivo'.

Por cierto, cabe notar que el conocimiento práctico no es conocimiento científico, lo cual no significa que aquél sea una especie de saber trascendental. En el caso del conocimiento práctico, no hay nada oscuro acerca de qué es lo que 'trasciende' en nuestro tiempo

a la teorización o a la formalización rigurosamente científicas. El motivo de que saber, por ejemplo, de *vinos* o de *cocina* no sea conocimiento científico consiste en que el criterio para determinar la excelencia de un platillo o de un vino es un paladar humano satisfecho y, en lo que a ello toca, no cualquier paladar. La cuestión de si la elaboración de una buena comida o un buen vino puede reducirse 'en principio' a la ciencia carece de interés, porque 'en principio' no guarda ninguna relación con la vida humana real en el futuro cercano. Sin duda, no existe una 'imposibilidad lógica' para que algún día no tengamos una teoría tan completa sobre nuestra propia naturaleza que programemos una computadora para determinar qué puede satisfacernos y qué no. Pero si esto llegara a suceder alguna vez, no podría dejar de alterar nuestra naturaleza. Si en un futuro nos volvemos tan transparentes para nosotros mismos que llegara a desaparecer la distinción entre conocimiento práctico y teórico, desaparecerán también sin duda instituciones como la ciencia, la filosofía y la literatura bajo su presente forma. Y si no es posible reducir a una 'ciencia' el cocinar bien o el fabricar un buen vino, ¡cuánto menos puede esperarse reducir a una 'ciencia' el *vivir* bien!

Pero el hecho de que no podamos hacer esta reducción no implica que reflexionar sobre cómo vivir bien no sea una empresa racional, o que no haya ningún conocimiento objetivo en ello. Por lo demás, la ciencia misma comprende tanto al conocimiento práctico como al teórico. Aunque las teorías científicas particulares constituyen el verdadero paradigma del conocimiento teórico, parece improbable que, en el futuro cercano, el saber cómo formular buenas teorías científicas se convierta en una ciencia teórica.

Quisiera subrayar otra vez este punto: mi respuesta a la interrogante de si la moral puede ser 'en principio' una ciencia es que esto sólo podría llegar a ocurrir, en todo caso, cuando la naturaleza humana resulte completa y científicamente transparente para los hombres. Puesto que ello no sucederá por lo menos en un milenio, si acaso, lo mejor será dejar de lado la cuestión de la moral como ciencia.

Una circunstancia planteada por Baker y Grice, que tiene importancia para mi presente tema de discusión, es que la tradición filosófica ha pasado por alto o menospreciado de forma sistemática el papel de la imaginación en el razonamiento práctico. Uno de sus ejemplos parece muy sencillo, pero por lo mismo enormemente instructivo. Un hombre escala una montaña. Se detiene a mitad del

camino, porque no está seguro por dónde seguir. Se imagina que toma una ruta. En su imaginación continúa subiendo hasta un punto determinado, y luego se topa con una dificultad para la que no encuentra, siempre en su mente, ninguna salida. Entonces se ve a sí mismo subiendo por un camino distinto. En esta ocasión se imagina que logra recorrer sin dificultades todo el trayecto hasta la cima. Así que toma la segunda ruta.

El punto es que ésta puede ser una manera perfectamente racional de resolver un problema práctico y, no obstante, se trata de una clase de razonamiento que no necesita ser reductible en ningún caso a un tipo de razonamiento lineal, expuesto proposición por proposición. El montañista funciona, por así decirlo, como una computadora analógica, en vez de digital, cuando resuelve su problema 'desarrollando' en su imaginación, de la manera más vívida posible, 'lo que sucedería exactamente si...'. Por supuesto, decir que esto no es un razonamiento proposicional lineal no equivale a negar que, *después* de que ha imaginado 'lo que sucedería si...,' pueda *expresar* en palabras las consideraciones pertinentes. En otros términos, no es preciso que el hombre haya *almacenado* la información pertinente *en forma de palabras*. Grice y Baker sugieren que imaginar modos de vida, o aspectos particulares de los modos de vida, posee una enorme importancia para la argumentación moral y, en general, para la argumentación práctica; y que no hay motivo para que la 'retórica' sea considerada, como normalmente sucede, un mero recurso propagandístico, sino que puede ser un instrumento legítimo para lograr que una persona imagine *vívidamente* cómo sería vivir de una manera en lugar de otra, o al menos puede ser un camino para hacer que perciba vívidamente cuál es el *atractivo* de una moral en comparación con el de otra.

Lo que me gusta del planteamiento de Grice y Baker es que sugiere que el razonamiento moral puede ser razonamiento en el sentido pleno del término, proponiendo al mismo tiempo que se trata de algo que no sólo implica a las facultades lógicas, en su acepción estricta, sino a nuestra capacidad íntegra para imaginar y sentir, en suma, nuestra sensibilidad total.

Quisiera hablar ahora de literatura. Durante mucho tiempo ésta se ha rehusado a ser, al menos primordialmente, propaganda de una moral, una filosofía o una ideología. En el Japón del siglo X, la señora Murisaki rechazaba en sus escritos la idea de que el propósito

de la literatura fuera la elevación moral, y lo hacía con la misma sinceridad de cualquier novelista de hoy. La literatura no obra igual que la retórica moral o ideológica. Normalmente, no nos ayuda a imaginar formas *ideales* de vida; no ofrece *soluciones*, o por lo menos no es frecuente. La novela, en especial, consigue auxiliarnos en la recreación imaginativa de los dilemas morales, en su sentido más amplio. (No estoy afirmando que esto constituya el valor estético de la novela; de hecho, no intento tratar en lo absoluto cuestiones de valor estético. Pero sostengo que es un rasgo importante del género, según se ha desarrollado en los últimos siglos.) En ocasiones, se dice que la literatura describe el 'predicamento humano', lo que quizás representa una forma de referirse a ello. Pero la pomposidad de la frase oscurece la esencia del problema. Lo importante no es que haya *un* predicamento que sea *el* predicamento humano, y que la literatura a veces lo describa; sino que, por *muchas* razones, parece que resulta cada vez más difícil imaginar *cualquier* modo de vida que sea a la par ideal y viable. Y la literatura nos enfrenta a menudo, con detalles emocionales extremadamente vívidos y ricos, con el porqué y el cómo de que esto parezca ser así en diferentes sociedades, en épocas distintas y desde diversas perspectivas. Me gustaría sugerir que si el razonamiento moral es, a nivel reflexivo, la crítica consciente de las formas de vida, entonces la apreciación sensible de predicamentos y dilemas que tiene lugar en la imaginación debe ser esencial para un razonamiento moral sensible. Es cierto que las novelas y las obras de teatro no nos ponen ante el conocimiento moral. Pero (frecuentemente) hacen por nosotros algo de lo que se debe hacer para que lo obtengamos.

En un mundo donde los libros revelados ya no gozan de aceptación general, muy poca gente negaría que el conocimiento científico debe desempeñar algún papel en la resolución de los problemas morales, tanto a nivel social como individual. Pero es fácil confundirse sobre la función que puede y debe tener la literatura. Al mismo tiempo, las ciencias sociales y psicológicas suelen decepcionarnos. De modo que propendemos a sentirnos, a la vez, desilusionados con respecto a la ciencia y desorientados con respecto a la situación de la literatura. Pienso que, mientras no tengamos una mejor perspectiva del razonamiento moral que la que nos ha servido de alimento últimamente, no seremos capaces de llegar a ningún tipo de consenso razonado sobre cómo pueden contribuir la literatura o las ciencias a nuestras preocupaciones profundas.

Aunque resulta comprensible que estemos desilusionados de la ciencia desde el punto de vista de los problemas morales y, más ampliamente, humanos y sociales, hay una objeción que he oído discutir entre algunos de mis colegas humanistas, que me parece fundamentalmente desencaminada.

En una ocasión me tocó escuchar a un grupo de profesores de humanidades del seminario Christian Gauss de Princeton cuando discutían acerca de los científicos. La opinión general era que estos últimos son zoquetes ignorantes que nunca leen nada de mayor valor literario que el último número de *Galaxy*. Varios oradores añadieron que los científicos son engreídos incluso fuera (*particularmente* fuera) de sus especialidades profesionales, y que resultan peligrosos a nivel social porque creen hallarse en una posición privilegiada para asesorar a los gobiernos, y porque los gobiernos se inclinan a dejarse impresionar por sus credenciales científicas y tienden a adoptar sus consejos sobre asuntos políticos. Esto último —la asesoría en asuntos políticos— recibió el calificativo de 'simplista'.

Los humanistas no veían nada extraño en calificarse *a sí mismos* como 'cautos', 'conscientes de la complejidad de las cosas', etc., aun cuando intelectuales *humanistas* del presente siglo se hubieran unido en importante número a las siguientes causas célebres: psicoanálisis, existencialismo (esto en cuanto a los intelectuales del continente) y algo vagamente descrito (por Paul Goodman, Norman Mailer, Kenneth Tynan *et al.* como 'revolución sexual'. ¿Qué científico digno de ese nombre ha estado dispuesto a afirmar categóricamente que *todos* los sueños son 'realizaciones de deseos'? ¿Que la vida humana carece de sentido, hasta que uno le da significado por un acto de 'compromiso' completamente arbitrario? ¿Que necesitamos una 'revolución sexual' absoluta (de la cual no existe ninguna descripción)? Los humanistas se hallan demasiado dispuestos a recordar que esos pronunciamientos son los de una minoría en sus filas y a olvidar que los descabellados pronunciamientos sobre cuestiones de política pública que provienen de los científicos corresponden a una minoría igualmente pequeña de esa comunidad. Es claro que tanto científicos como humanistas han sido a veces culpables de apoyar a distintos tipos de discriminación, colonialismo, etc., y que ambos han defendido a la decencia.

Por fortuna, para mis propósitos actuales, no es preciso discutir la mayoría de estos 'casos', porque en realidad se componen de opiniones prejuiciadas acerca de los científicos, las cuales no resultarían

muy pertinentes con respecto al valor o la naturaleza de la ciencia aun cuando fueran ciertas. Lo interesante es la pretensión de que, por así decirlo, Eurípides, Freud y Dostoievski son 'todo lo que vosotros conoceis sobre este mundo y todo lo que necesitais conocer'.

Parece difícil saber qué se puede decir acerca de esto porque, dejando a un lado el escabroso problema sobre la legitimidad del psicoanálisis, hay que admitir que *Medea* y *Los hermanos Karamazov* constituyen obras de arte grandiosas y conmovedoras. Somos ante todo seres humanos, y estas obras nos conmueven en tanto que somos seres humanos. ¿Pero por qué la concesión de un alto valor al arte es incompatible con la concesión de un alto valor a la ciencia y también a otras cosas buenas?

Nos acercamos al problema crucial si advertimos que de acuerdo con algunos —con los pensadores que quiero criticar— el valor de, digamos, Dostoievski, no es puramente estético en el sentido que lo tiene, o a veces se supone que lo tiene, una pintura abstracta. Estos pensadores creen que los dramaturgos griegos, la psicología freudiana y la novela rusa encarnan *conocimiento —conocimiento sobre el hombre*. De este modo, entran y no entran en conflicto con la ciencia. Están en conflicto con ella porque representan un tipo de conocimiento rival y, por lo tanto, le disputan su pretensión de monopolizar el conocimiento confiable. Pero se trata de un *tipo* de conocimiento rival y, en consecuencia, no es suceptible de comprobación científica. Si añadimos la pretensión de que este tipo de conocimiento rival es en cierto modo 'superior' o más importante que el científico, nos encontramos frente a una verdadera posición oscurantista; sin duda no la que sostiene el estudioso o el crítico literario serio, sino la correspondiente a la religión de la literatura.

No quiero sugerir que quienes proponen esta postura no estén detrás de algo, sino que la discusión sobre el tema se complica en seguida con cuestiones espinosas y difíciles. Ofrezco con cierta modestia las siguientes observaciones acerca de la forma en que, digamos, la novela encarna o no encarna un 'conocimiento sobre el hombre'.

Afirmar que las novelas proporcionan un conocimiento sobre el hombre me parece tan equivocado como decir categóricamente que no lo hacen. La situación es más compleja de lo que sugeriría una simple aseveración en favor de cualquier postura. No importa cuán profundas parezcan las observaciones psicológicas de un novelista, no pueden llamarse *conocimiento* si no han sido probadas. Aseverar

que un lector perceptivo es capaz sencillamente de *darse cuenta* que esas observaciones psicológicas no son sólo plausibles, sino que contienen una especie de verdad universal, equivale a regresar a la idea del conocimiento por intuición sobre asuntos empíricos, al 'método de lo que es conforme a la razón'. Para tomar algunos ejemplos de novelas menos difíciles que las de Dostoievski: si leo *Viaje al fin de la noche* de Celine, no *aprendo* que el amor no existe, que los seres humanos son odiosos y odiables (aun si —y estoy seguro de que no es así— esas proposiciones fueran verdaderas). Sino que aprendo a ver el mundo como se le aparece a alguien que está convencido de que esa hipótesis es correcta. Veo la verosimilitud que posee esa hipótesis, lo que sucedería si *fuera* verdadera, cómo hay una persona capaz de pensar que lo *es*. Pero todo esto no es todavía conocimiento empírico. Y sin embargo, no es correcto decir que no es conocimiento en absoluto, porque la conciencia acerca de una nueva interpretación de los hechos, aunque resulte repulsiva, de una construcción que, según me doy cuenta ahora, puede erigirse sobre los hechos, no importa con cuánta malignidad, es un tipo de conocimiento. Es el conocimiento de una posibilidad. Es conocimiento *conceptual*.

Quizás parezca extraño que califique de conocimiento *conceptual* a algo tan real y 'empírico' como una visión del comportamiento de los seres humanos y de qué los 'hace palpitar'; pero no es más que esto, a menos que sea comprobado, si no científicamente, siquiera por la experiencia real de hombres y mujeres inteligentes y sensibles. Pensar en una hipótesis que no se haya considerado antes constituye un descubrimiento *conceptual*, y no empírico, aunque puede llevar a esto último si la hipótesis resulta ser correcta. Sin embargo, el 'conocimiento de una posibilidad', que nos proporciona la literatura, no debe referirse a una *mera* posibilidad. Que la posibilidad ofrecida por Celine *sea* una 'mera' posibilidad constituye, después de todo, una de las razones por las que no le damos una calificación mayor como novelista. Así, la situación se complica de nuevo; hay tantos elementos conceptuales como empíricos en el conocimiento que obtenemos de la literatura.

Pero podría objetarse que la 'hipótesis' del novelista no está sujeta a comprobación científica. No se trata de esa *clase* de hipótesis. Esto es cierto con frecuencia, pero no siempre. Si Celine estuviera en lo correcto, si todas las instancias aparentes del amor fueran en realidad ejemplos disfrazados del interés más egoísta por uno mismo, no

me cabe duda de que podrían encontrarse muy diversas evidencias empíricas de esto. Quizás la dificultad en este punto estriba en que los operacionalistas han convencido a demasiada gente de que todas las hipótesis científicas llevan su comprobación consigo, y por lo tanto, que cualquier enunciado cuya comprobación no es obvia no puede tener ningún lugar dentro de la esfera científica. Pero esto es una burda caricatura de la ciencia, aun cuando muchos científicos han aprendido a creer en ella. Con mucha frecuencia, el genio de un científico radica precisamente en imaginar alguna manera de probar un enunciado que no parecía demostrable en absoluto.

Para tomar un ejemplo distinto y más creíble que el de Celine: una de las virtudes de *El cuaderno dorado* consiste en que presenta una descripción extremadamente verosímil de cómo era ser comunista en la década de los cuarenta. Sin embargo, no puede afirmarse que después de leerlo se haya adquirido el *conocimiento* de cómo era ser comunista en la década de los cuarenta, a menos que se sepa por una fuente independiente que la descripción de Doris Lessing es verdadera en los hechos. Es posible sentirse convencido al leer *El cuaderno dorado*, uno puede decirse a sí mismo que *así debió ser*; pero a menos que se quiera sustituir la correspondencia con los hechos objetivos y la exposición de datos adecuados que respondan al criterio de 'conocimiento' por la credibilidad subjetiva y la concordancia con 'lo que es conforme a la razón', no es lícito aseverar: '*sé* que así era'. Uno no *sabe*, y apenas a la siguiente semana puede sentirse convencido, por una novela igualmente verosímil, de que todo debió ocurrir de una manera por completo distinta a la descripción de Doris Lessing.

No obstante, Doris Lessing nos presenta algo muy importante en *El cuaderno dorado*, cuyo valor no depende de que su descripción del Partido Comunista Británico sea correcta o incorrecta (aunque sí de la exactitud de su descripción sobre cómo era considerado ese partido y de lo acertado de una parte de su crítica a los comunistas, por ejemplo, que eran totalitarios y manipuladores). La autora nos presenta un cierto dilema moral, un problema o un conjunto de problemas morales, algunos de los cuales se relacionan con el hecho de ser una mujer en el presente siglo, otros con ser una persona que posee conciencia social en el presente siglo, de modo que el dilema moral pudo ser vivido por un individuo perfectamente posible en un periodo perfectamente definido. Lo que estoy sugiriendo es que si queremos reflexionar de manera racional sobre el feminismo,

el comunismo, el liberalismo o simplemente la vida del siglo XX, lo que Doris Lessing ofrece a nuestra sensibilidad tiene una enorme importancia.

Considerar que la novela en sí misma nos presenta una especie de conocimiento no científico sobre el hombre equivale de algún modo a convertirla en algo demasiado parecido a las *proposiciones*. Dije al principio que hay una interpretación según la cual existe una suerte de 'conocimiento no científico del hombre'; aunque la moral no es un 'conocimiento no científico del hombre' según la acepción que le dan las personas a las que recién aludimos, constituye un conocimiento práctico acerca de cómo vivir; y éste no es 'científico' en el sentido de ser conocimiento teórico. Más aún, el conocimiento práctico, como ya vimos, involucra a nuestra capacidad plena para sentir e imaginar. Creo que las personas a las que me referí estaban en lo cierto con respecto a que la literatura se vincula de algún modo con una clase de conocimiento cercano al núcleo de la preocupación moral, el cual no constituye un 'conocimiento científico' desde ningún punto de vista que sea razonablemente convencional. Pero decir que la literatura 'proporciona' un conocimiento no científico del hombre simplifica demasiado las cosas.

Así como algunos pensadores tienden a considerar la literatura como conocimiento, otros suelen contemplar en el presente a la ciencia como filosofía. Esta última tendencia resulta comprensible. Muchas cuestiones filosóficas tradicionales —por ejemplo, la infinitud del espacio y del tiempo, y la naturaleza del espacio, el tiempo y la materia— han sido relegadas a la física por los filósofos de mentalidad más realista. Cuando mucho, el filósofo profesional espera hallarse en aptitudes de aclarar los resultados que obtiene el físico. Las ciencias sociales se encuentran, por muchas razones, en un estado menos sólido que las ciencias físicas, pero ya existe la esperanza, si no el logro, de que las ciencias sociales y psicológicas puedan algún día arrojar verdadera luz sobre la naturaleza de la sociedad y el hombre. Sin duda, hoy son pocos los filósofos de cualquier escuela que se atreverían a escribir sobre la justicia sin considerar los frutos de la teoría económica, o sobre las partes de la psique sin considerar la teoría psicológica. Pero si la ciencia es una filosofía, adolece del hecho de que se compone por completo de metafísica y no tiene nada de ética; y la metafísica es ciega sin la ética. La ciencia social a menudo parece estéril precisamente porque se interesa en la mera des-

cripción de lo que tenemos como sociedad ahora, y no en la posibilidad de una sociedad al mismo tiempo viable y justa; y la ciencia psicológica rara vez discurre sobre el hombre en sociedad, o sobre la existencia posible, en *cualquier* sociedad, de modos de vida que sean simultáneamente morales, viables y gratificantes.

No estoy exigiendo que se construya ciencia sólo con fines prácticos o sólo para ilustración moral. Por supuesto, el conocimiento en sí mismo es y debe ser un 'valor terminal' para hombres y mujeres cultos. Pero sostengo que resulta difícil comprender sin distorsión incluso la significación *filosófica* de la ciencia, ya no digamos su significación práctica, cuando la ciencia y la reflexión moral se hallan separadas de un modo tan tajante como lo han llegado a estar en nuestra cultura.

La teoría de la relatividad tiene importancia sin duda para la naturaleza de la materia, el espacio y el tiempo; pero también afecta, por ejemplo, a la naturaleza entera y a la posibilidad de la certeza humana. El descrédito de la geometría euclideana no fue *sólo* el descrédito de una teoría del espacio; representaba el paradigma de la certeza obtenida a través de un razonamiento *a priori* y, aún más, el paradigma que tanto Platón como Spinoza ofrecían al filósofo moral. La comprensión del papel que ha tenido el ideal de certeza en nuestra vida intelectual y moral durante más de dos mil años y la consideración cuidadosa de las implicaciones que ha tenido su desaparición para el razonamiento moral (y social y religioso), así como para la ciencia pura, no son incompatibles con la apreciación de la relatividad 'en sí misma'; de hecho, son inseparables para lograr esto último con propiedad. Y gran parte del vacío que se observa en la ciencia social del presente surge de la tentativa de estudiar cuestiones sociales y psicológicas con un ideal de 'objetividad' completamente falso, el cual pierde incluso las conexiones de las ciencias sociales entre sí, además de pasar por alto los problemas de mayor importancia para la reflexión moral.

Sin embargo, cabe una advertencia. Renunciar a los falsos ideales de la 'objetividad' no significa que deba asumirse el punto de vista de que toda ciencia social tiene que sacar al mundo del caos en que se encuentre durante, digamos, los próximos cien años. A pesar de su importancia, esto se halla muy lejos de la interrogante fundamental, la cual no consiste en cómo hacer reformas individuales —aunque sean tan grandes como el 'socialismo'—, sino *hacia dónde* queremos ir.

Trataré de resumir estas reflexiones tan diversas. En este artículo no he intentado demostrar que las teorías emotivas del discurso ético sean erróneas o que lo sea el relativismo ético. Ello es tema para ensayos filosóficos técnicos. He indicado que, en mi opinión, esas perspectivas filosóficas de moda, las cuales se hallan extensamente difundidas en la actualidad, están equivocadas y que al menos algunos de los principios éticos poseen probablemente una gran medida de objetividad. He sugerido, siguiendo el camino de Judy Baker y Paul Grice, que la 'objetividad' de los principios éticos o, en términos más amplios, de las 'morales', está relacionada con cosas como la magnitud de su atractivo, la capacidad para oponer resistencia a ciertas clases de crítica racional (que no he tratado de enumerar), la viabilidad, la idealidad y, por supuesto, con lo que *se siente* en realidad al vivir o tratar de vivir a su amparo. También he señalado que el conocimiento moral es lo que los filósofos llaman 'conocimiento práctico' en oposición al conocimiento teórico, y que la imaginación y la sensibilidad constituyen *instrumentos* esenciales del razonamiento práctico. Mi propósito no ha sido defender aquí esos puntos de vista, sino ver cuál es su importancia, en caso de ser correctos, no para la filosofía sino para la literatura y la ciencia.

He sostenido que la idea de que la moral es sólo 'subjetiva', y su corolario de que el razonamiento ético sirve únicamente como instrumento (relacionado con la selección de los medios para fines escogidos *de manera arbitraria*) o es una contradicción en los términos, resulta terriblemente destructiva no sólo para la moral misma sino para toda la cultura. La única forma en que podemos entender, por un lado, cómo nos ayuda realmente la imaginación literaria a comprendernos a nosotros mismos y a la vida, y por el otro, cómo se relaciona la ciencia con los problemas metafísicos, es mediante una percepción adecuada del razonamiento moral, término con el cual no quiero significar solamente a la reflexión sobre el deber o la virtud, sino que me refiero al razonamiento moral en su sentido más amplio, al razonamiento sobre cómo vivir.

TERCERA PARTE

REFERENCIA Y COMPRENSIÓN

REFERENCIA Y COMPRENSIÓN

Mi tesis es que la teoría de la comprensión del lenguaje y la teoría de la referencia y la verdad tienen mucho menos que ver entre sí de lo que gran parte de los filósofos ha supuesto. Quiero explicar por qué. Creo que si estoy en lo correcto, esto arrojará luz sobre una buena cantidad de cuestiones filosóficas; por ejemplo, por qué Wittgenstein tuvo una teoría del significado pictórica y otra 'utilitaria' (por supuesto, en diferentes etapas), y qué ocurría en el siglo XIX con las objeciones a una teoría de la verdad por correspondencia.

Comprensión del lenguaje

Me parece que la explicación de que comprender un lenguaje *consiste* en ser capaz de usarlo (o de traducirlo a un lenguaje que uno *pueda* emplear) es la única que existe por ahora. Quizás Michael Dummett logre desarrollar una explicación alternativa (sé que quiere hacerlo), pero por el momento no conozco ninguna. En segundo lugar, no creo que deba pensarse que la 'capacidad de usar' un idioma proviene del aprendizaje de pequeñas escenas separadas del tipo que Wittgenstein utiliza en las primeras páginas de sus *Investigaciones* para ilustrar la noción sobre el 'juego del lenguaje'. Algunos de sus seguidores parecen entender el idioma de esa forma, como consistente en 'usos' *desconectados* (por ejemplo, la expresión 'ése es un juego de lenguaje diferente' que en ocasiones emplean tales sujetos), pero no creo que Wittgenstein sea culpable de esto; y, en todo caso, no es esencial para su doctrina.

Aunque no cabe esperar un modelo *conforme a la realidad* para el uso global del lenguaje, la obra de Carnap y Reichenbach contiene un modelo *demasiado simplificado* (para el lenguaje declarativo). Se trata del modelo del hablante/oyente que posee una lógica inductiva

(por ejemplo, una probabilidad métrica subjetiva, aunque no creo que ésta sea una buena forma de ver a la inducción), una lógica deductiva, un ordenamiento preferencial (aunque me parece que no es más que una forma idealizada de modelar las estructuras preferenciales del ser humano) y una regla de acción (por ejemplo, 'aumentar al máximo la utilidad estimada', si bien creo que en muchas situaciones esta regla resulta deficiente). Imaginemos una comunidad de tales hablantes/oyentes que acepta oraciones expresadas por otros (o les concede una alta probabilidad) y que está autorizada a formular oraciones cuya probabilidad sobrepasa un cierto valor. Incluso en este modelo tan excesivamente simplificado, el lenguaje afectará la conducta en una amplia variedad de formas. Y cuanto mejor sea la lógica inductiva y la deductiva, cuanto más realista sea la función de utilidad, la conducta de esas criaturas se parecerá más a la 'comprensión de un lenguaje'.

Dicho modelo no está atado a una concepción individualista del conocimiento o del lenguaje. Puesto que los hablantes pueden adquirir conocimiento unos de otros (el conocimiento es una 'enfermedad contagiosa', según la feliz frase de John McDowell), no es necesario que cada hablante que tenga una palabra en su vocabulario, por ejemplo, 'oro', sea capaz de decir si algo es o no es oro. De hecho, en tanto que cada uno de los que cuentan con la palabra 'oro' en su vocabulario posee una *cantidad mínima de información convencional* sobre el oro, puede participar en las discusiones colectivas sobre este último. Así, la característica que he llamado 'la división lingüística del trabajo' (en "The Meaning of 'Meaning'")[1] es perfectamente compatible con tal modelo, el cual asimila de inmediato también a otra idea que postulé en ese artículo —la importancia de los 'estereotipos' concebidos como cantidades mínimas de información asociadas a palabras. Finalmente, el modelo del hablante/oyente es *holístico* en esencia, es decir, las condiciones en las cuales se emite cualquier oración en particular y la conducta que resulta cuando se expresa una oración dada no dependen de una única cosa que pueda ser denominada el 'sentido' de la oración, sino del sistema completo. Cambiar la lógica inductiva, o la deductiva, o la función de utilidad afectará hasta cierto punto las condiciones de emisión y las respuestas de la conducta para cada oración del

[1] *Cfr.* mi *Mind, Language, and Reality; Philosophical Papers*, vol. 2, Cambridge University Press, 1975, cap. 12.

lenguaje (aunque, desde luego, los enunciados comunes sobre macroobservables estarán asociados con las experiencias canónicas de 'verificación' y 'falsificación'). Sin embargo, no toda modificación del 'uso' (en el sentido del sistema o en el de las condiciones de emisión para responder a oraciones con oraciones/conductas) es un cambio de *significado*. Desde mi punto de vista, el significado es una tosca cuadrícula puesta sobre el uso. Creo que existen criterios diferentes para afirmar que se ha producido una alteración en el significado para distintas clases de palabras; y para algunas (por ejemplo, los términos lógicos) la modificación de la dicotomía 'significado/cambio de la teoría' no posee ningún sentido. (El hecho de que nuestro criterio para el 'cambio de significado' sea tan variado e inasible se explica por las diferentes clases de intereses que poseemos con respecto a diversos temas y actividades. En este sentido, el lenguaje *es* una 'mezcla', aun si ésta pudiera ser representada de manera uniforme en algún nivel de abstracción.)

Lo anterior no es más que un esbozo para una explicación sobre el 'uso' de la comprensión, pero doy por sentado que este tipo de explicaciones resultan bastante familiares (excepto, quizás, la idea del significado como una tosca cuadrícula sobrepuesta al uso). Quisiera ahora pasar a los temas del realismo, la referencia y la verdad.

Realismo y correspondencia

En ninguna parte de esta explicación acerca del 'uso' se menciona *algo* sobre la correspondencia entre las palabras y las cosas, o las oraciones y los estados de hecho. Pero de ello no se sigue que tal correspondencia no exista. Algunos instrumentos presentan la particularidad de que las instrucciones de uso no incluyen ninguna *explicación* sobre el éxito de su empleo. Por ejemplo, las instrucciones para prender o apagar la luz —'sólo oprima el interruptor'— no aluden a la *electricidad*. Pero la explicación sobre el éxito de la utilización del interruptor como método para hacer que se prendan y apaguen las luces menciona sin duda a la electricidad. En mi opinión es en este sentido que la referencia y la verdad se relacionan menos con la comprensión del lenguaje de lo que los filósofos suelen suponer.

En el modelo antes bosquejado se puede emplear el idioma propio, por lo menos a nivel de un 'lenguaje objeto', sin ninguna noción

de verdad complicada. Por supuesto, se debe poder asentir y disentir, pero ello es posible en el modelo de Carnap-Reichenbach. En este, las instrucciones que la mente sigue no presuponen nociones como la de 'verdadero', sino que son instrucciones para asignar un gran peso a determinadas oraciones cuando se tienen ciertas experiencias, instrucciones para expresarlas, para llevar a cabo transformaciones sintácticas, para producir una conducta no verbal (por ejemplo, "levanta un brazo cuando estimes que 'levantar un brazo posee la máxima utilidad prevista'"), etc. Pero el *éxito* del 'programa para la utilización del lenguaje' bien puede depender de que exista una correspondencia adecuada entre sus palabras y las cosas y entre sus oraciones y los estados de hecho. Las nociones de verdad y referencia pueden resultar de gran importancia para explicar la relación del lenguaje con el mundo, sin que por ello sean tan cruciales para la teoría del significado (en el sentido de la teoría sobre la comprensión del lenguaje) como lo son, por ejemplo, para las teorías que *identifican* la comprensión con el conocimiento de las condiciones de verdad.

En esto pensaba cuando me referí a las dos teorías del significado de Wittgenstein: la teoría pictórica *era* incorrecta como teoría de la comprensión, por razones que Wittgenstein mismo detectó muy bien; pero no era completamente incorrecta en tanto teoría del funcionamiento del lenguaje. Los realistas objetan que resulta esencial considerar nuestras teorías como una especie de 'mapa' del mundo para que podamos explicar la forma en que nos ayudan, según es el caso, a guiar nuestra conducta. Pero la teoría 'utilitaria' *también* es correcta en tanto explicación de cómo *comprendemos* el lenguaje. Y no existe incompatibilidad entre ambos puntos de vista: después de todo, un mapa es lo que es en virtud de que *se lo utiliza* de cierto modo, lo cual no contradice sino que complementa el hecho de que el mapa sólo sirva si corresponde adecuadamente a una sección particular de la tierra o a lo que fuere. Hablar de uso y hablar de referencia forman parte de una sola explicación, igual que hablar de oprimir interruptores y de la electricidad que fluye por los cables.

Éxito

He empleado la noción de éxito. Intentaré ahora desarrollarla. Lo que en general 'tiene éxito' o 'fracasa' no es la conducta lingüística en sí misma sino la conducta total. Por ejemplo, decimos ciertas

cosas, vinculamos unos razonamientos con otros, manipulamos materiales de determinada forma y finalmente tenemos un puente que nos permite cruzar un río que antes no se podía atravesar. Y nuestro razonamiento y nuestra discusión forman parte del complejo total y organizado de la conducta en la misma medida que la elevación de pilotes de acero efectuada con nuestra grúa. Por lo tanto, no habría que hablar en realidad del éxito o el fracaso de nuestra conducta lingüística, sino de la *contribución* que éste hace al éxito de nuestra conducta total.

Hasta cierto punto, todas las posturas metafísicas proporcionan la misma explicación acerca de esto, es decir, de la contribución de la conducta lingüística al éxito de la conducta total, pero sólo hasta cierto punto. La explicación consiste en que algunas de las clases de creencias que sostenemos suelen ser *verdaderas* (o cualquier predicado que la tendencia en cuestión sustituya por 'verdadero', por ejemplo, 'justificadamente aseverable'). Algunos filósofos han sido tan poco precavidos que llegaron a proponer la máxima de que 'la mayoría de las creencias de un hablante son verdaderas' como una especie de principio *a priori* que rige a la traducción radical; pero me parece que es ir demasiado lejos. (En primer lugar, no sé cómo *contar* las creencias, así que no entiendo qué significa hablar de la mayoría de las creencias de un hablante. Y en segundo lugar, la mayoría de las creencias que las personas tienen sobre *algunos* temas —por ejemplo, la filosofía— son probablemente *falsas*.) Pero la mayoría de la gente tiene creencias verdaderas sobre dónde vive, cómo es su barrio, cómo llegar de un lugar a otro, etc. Y posee muchas creencias verdaderas sobre la manera en que reaccionarán los demás en distintas circunstancias y sobre cómo hacer ciertas cosas. Y *toda* posición filosófica admite a grandes rasgos el siguiente recuento:

(1) Por lo general, la gente actuará de un modo tal que alcanzará sus metas (tanto como sea posible en su situación dada), o que no se frustrarán sus expectativas o, en términos más concretos, que obtendrá alimentos, alojamiento, etc., encontrará amigos y compañeros, etc., logrará trasladarse de un lugar a otro, evitará los peligros, etc., *si* sus creencias son verdaderas.

(2) Muchas creencias (de los tipos que mencioné y de otros tipos pertinentes) *son* verdaderas. (Aunque algunas son falsas, lo cual forma parte de la explicación de los *fracasos* que sufren las personas.)

(3) Por lo tanto, como consecuencia de (1) y (2), la gente tiende a alcanzar ciertas clases de objetivos.

Me parece que esto no es controvertible. Toda explicación acerca de la verdad intenta hacer que (1) y (2) sean correctas, y así permitirnos explicar (3) a través de ellas. Por ejemplo, la explicación de un pragmatista o de un fenomenólogo interpreta 'verdadero' de tal forma que un enunciado verdadero corresponde a las expectativas condicionales que se cumplirán siempre que se lleve a cabo la acción que es la condición antecedente de la expectativa condicional. De esta manera, se observa (1), a saber, si uno actúa siguiendo creencias *verdaderas*, sus expectativas no se verán frustradas. Y todas las posiciones filosóficas (idealismo, pragmatismo, realismo, etc. —excepto el escepticismo extremo—) sostienen que la mayoría de las creencias que posee la gente sobre dónde vive, etc., son verdaderas. Así, se acepta (2). Por lo tanto, para todas las tendencias, la 'verdad' desempeña el mismo papel en lo tocante a explicar la contribución que realiza la conducta lingüística al éxito de la conducta total *al menos hasta este punto:* que (3) se explica a través de (1) y (2).

Sin embargo, éstas no son todas las condiciones deseables que uno quisiera imponer en una explicación de la verdad. Por ejemplo, en otro lugar[2] he argumentado que algunos sustitutos positivistas para la noción de verdad (por ejemplo, 'es simple y lleva a predicciones verdaderas') carecen de la propiedad de que el conjunto de las teorías aceptables (las que tienen la propiedad en cuestión) sea aceptable. Por lo tanto, la lógica deductiva no preserva la aceptabilidad.[3] Pero los científicos consideran como admisibles las *consecuencias lógicas* de las teorías aceptables. (En el sentido de que si la consecuencia es inadmisible, hay que volver atrás y revisar la decisión de que *todas* las premisas eran aceptables.) Así, una explicación de este tipo, la que dice que estamos buscando una especie de aceptabilidad que carece de la propiedad de conclusión deductiva, no logra justi-

[2] En *ibid.*, cap. 11.

[3] Sin embargo, la lógica *no clásica* puede preservar a un cierto tipo de 'aseverabilidad justificada'. La posibilidad de remplazar la noción de verdad realista por una noción de aseverabilidad justificada, interpretando la lógica del lenguaje natural en términos que no sean los clásicos, fue propuesta por Michael Dummett; está más o menos insinuada en su famoso artículo sobre la verdad ('Truth', *Proceedings of the Aristotelian Society*, vol. 59 (1958-1959), pp. 141-162) y yo la traté en "What is 'Realism'?" (*Proceedings of the Aristotelian Society* (1975-1976), pp. 177-194).

ficar las normas de la práctica científica. Es precisamente a causa de que existen otras condiciones deseables que una explicación acerca de la verdad debe satisfacer, además de dar cuenta mínimamente de la contribución aportada por la conducta lingüística al éxito de la conducta total (lo que he significado con (1), (2) ∴ (3)), por lo que hay motivos para rechazar como inadecuadas algunas posiciones metafísicas sobre la naturaleza de la verdad y/o la aceptabilidad.

Además de que nos proporciona la explicación mínima sobre la contribución realizada por la conducta lingüística al éxito de la conducta total, y de que nos presenta el hecho de que la verdad (o cualquier noción de aceptabilidad que uno proponga como sustituto si por alguna causa se objeta la noción de verdad) se preserva en virtud de las reglas (apropiadas)[4] de la lógica deductiva, existe en particular otra condición deseable que una explicación satisfactoria de la verdad debería cumplir: tendría que *explicar la confiabilidad de nuestro aprendizaje*. Me permitiré ahora explayarme sobre esto último.

Explicación realista de la confiabilidad epistémica

La explicación mínima sobre la contribución de la conducta lingüística al éxito de la conducta total recurre a la idea de que la mayoría (o al menos muchas) de nuestras creencias sobre ciertos asuntos son verdaderas. Pero tales creencias no constituyen un acervo definitivo que se fija de una vez para siempre. Cambian constantemente como resultado del aprendizaje (y utilizo el término en un sentido amplio que incluye a la percepción). De esta manera, lo afirmado por la premisa (2) de la explicación mínima es que ciertos tipos de aprendizaje son *confiables*, con lo que quiero significar que conducen a un gran número de creencias verdaderas. ¿Cómo explicar esta confiabilidad?

Desde luego, el tipo de 'explicación' que uno cree que debe darse dependerá de las características de la posición filosófica personal. Los realistas contemporáneos a menudo difieren de los realistas de otros siglos en que rechazan la noción de verdad *a priori*, o al menos en que no ponen mucho 'peso' en ella. Asimismo —aquí me

[4] Creo que es posible decidir que la lógica clásica es incorrecta por razones realistas más que intuicionistas (*cfr.* más adelante la nota 7). Pero en este caso el teórico debe mostrar que la lógica correcta (de acuerdo con la nueva teoría) tiene a la lógica clásica como un 'caso especial' en muchas situaciones, de tal modo que podamos contar con el uso apropiado de la lógica clásica *donde ésta tenga éxito*.

arriesgo de nuevo a hacer generalizaciones sobre filósofos que tienen una gran cantidad de desacuerdos entre ellos— suelen mostrarse escépticos con respecto a la idea de un 'método científico' fijo e inalterable. Y si no se consideran como dados *a priori* los métodos para la obtención de conocimiento, y en cambio se considera que están inmersos en la historia y evolucionan con ella, parece inverosímil que se vea como defendible ninguna dicotomía profunda entre el *método* y el *contenido* del conocimiento.[5] Lo que nuestros 'métodos' son en cualquier campo depende de cuáles sean nuestras *creencias* sobre la materia de estudio de ese campo y de cuáles nuestras creencias en otros dominios. Desde este punto de vista, el problema de explicar la confiabilidad del aprendizaje no consiste en demostrar *a priori* que nuestros métodos de aprendizaje son confiables, sino que puede comprenderse esa confiabilidad como un hecho natural, que se la puede explicar, o aproximarse a su explicación, como se explican o se avanza hacia la explicación de otros hechos naturales. Esta posibilidad puede contemplarse como una especie de 'control de la consistencia' (en un sentido informal de 'consistencia') para la concepción metafísica como un todo.

Para entender la manera en que se lleva a cabo este plan, tomemos el sencillo caso de la percepción visual: un hablante normal mira un objeto, digamos una alfombra, y dice: 'la alfombra es verde' o 'la alfombra no es verde'. ¿Es confiable?

Parte del asunto estriba en la explicación causal común acerca de la visión. Lo que ésta nos proporciona, hasta donde ha sido desarrollada (y nadie negará el enorme progreso realizado en la comprensión de la visión a color), es una explicación de los siguientes hechos:

(4) Si la alfombra es verde, el hablante probablemente aceptará 'la alfombra es verde'.

(5) Si la alfombra no es verde, el hablante probablemente aceptará 'la alfombra no es verde'.

Pero lo que queremos es:

(6) El hablante probablemente acepte cualquier enunciado que sea *verdadero*.

[5] Una interesante defensa de esta versión del realismo aparece en Richard Boyd, 'A Causal Theory of Evidence', *Nous*, vol. 7 (1973), pp. 1-12.

Ahora bien, el realista admite asimismo una definición convencional de verdad[6] para el lenguaje. Dicha definición establece una correspondencia entre las cosas y las palabras (por ejemplo, *'la alfombra' se refiere a una alfombra definida contextualmente; 'es verde' se refiere a cosas verdes: oraciones del tipo 'Las NV son verdaderas si y sólo si el objeto al que se refiere 'La N' es el referido por 'V'*), y presenta las siguientes características:

(a) Satisface el criterio *T* de Tarski, por ejemplo
'La alfombra es verde' es verdadera si y sólo si la alfombra es verde,

y

(b) El término 'verdadero' se intercambia con funciones de verdad y con operadores como 'probablemente'.

En virtud de estas características, (4) y (5) implican:

(4′) Si la alfombra es verde, el hablante probablemente aceptará el enunciado que sea verdadero del par 'la alfombra es verde', 'la alfombra no es verde'.

(5′) Si la alfombra no es verde, el hablante probablemente aceptará el enunciado (del mismo par) que sea verdadero.

Y puesto que (en la situación considerada) la alfombra es verde o no verde:

(6′) El hablante probablemente aceptará el enunciado que sea verdadero (del par en cuestión en la situación considerada).

El enunciado (6′) dice que una cierta forma de aprendizaje (la percepción visual en el caso de una alfombra de color uniforme) es confiable. Y hemos delineado el modo en que un realista puede dar cuenta de esta confiabilidad *desde el interior* de nuestro sistema conceptual total (la teoría causal de la percepción y del uso del lenguaje más la teoría semántica de la verdad), mientras reconstruye ese sistema conceptual. De esta manera, el 'control de la consistencia' resulta satisfactorio. (Por supuesto, todo lo anterior es programático, pero también lo es nuestra explicación de la mayoría de los hechos empíricos.) Para el caso de las formas más complejas de aprendizaje, la explicación causal de la confiabilidad es, desde luego, mucho más programática. Pero a menos que queramos desechar o ignorar

[6] Utilizo el término 'definición de verdad' de acuerdo con la semántica tarskiana común. Una variante interesante ha sido propuesta hace poco por Saul Kripke en 'Outline of a Theory of Truth', *Journal of Philosophy*, vol. LXXII, núm. 19, 6 de noviembre de 1975, pp. 690–715. Esta variante permite que se hable de *referencia* y de *verdad* en el *lenguaje objeto mismo*, al costo de modificar el criterio *T* de Tarski.

el cuerpo completo de nuestra ciencia natural y nuestra especulación científica, no existe en el campo ninguna descripción alternativa que sea homogénea en el presente.

Un caso interesante es el de la inferencia para una *teoría*. En esto, como muchos lo han subrayado, parece necesario presuponer que tenemos un 'conocimiento *a priori* deficiente', en el sentido de que se desarrolló mediante un 'ordenamiento por simplicidad' de las teorías o una 'probabilidad métrica previa', lo cual no resulta 'tan mal' por cuanto no asigna una probabilidad irremediablemente baja a las teorías que son verdaderas en el mundo real. No podemos explicar cómo se produjo esto, pero al menos el hecho de que haya ocurrido no es *incoherente* con nuestras explicaciones actuales sobre el desarrollo de las especies, y sin duda se llevará a cabo una labor mucho mayor tanto en lo que respecta a la historia de ese desarrollo como a la estructura de nuestra capacidad de inferencia. El desorden de la situación no representa una fuente de desaliento para el realista de orientación científica y antiapriorístico; es exactamente lo que esperaría en el presente estado de la investigación sobre un problema complejo que toca los campos de la biología, la psicología y la lógica inductiva.

No resulta difícil entender el papel que desempeña en esto la idea de la *correspondencia* entre casos de lenguaje y realidad extralingüística. La teoría causal de la confiabilidad nos dice que, por ejemplo, cuando se produce una cierta situación (que la alfombra sea verde) el hablante emite una determinada oración ('la alfombra es verde'). La teoría semántica de la verdad nos indica que la oración es *verdadera* sólo en caso de que se produzca esa situación —la correspondencia implicada en el argumento causal es igual a la establecida por la definición de verdad. Por lo tanto, se nos podría concebir como sistemas que emiten puntualmente oraciones verdaderas cuando se produce una cierta gama de situaciones. El presupuesto (2) de la explicación mínima sobre la contribución de la conducta lingüística al éxito de la conducta total queda *explicado* (en líneas generales y de un modo programático).

Otra condición deseable para una explicación sobre la *verdad* consiste en que se dé razón (en el mismo sentido) de la exactitud del supuesto (I). Para la mayoría de los *idealistas* esto no representa ningún problema; un enunciado 'verdadero' *es* el que se corresponde con el tipo correcto de expectativas acerca de las consecuencias de la conducta. Pero el idealista también se encuentra con obstáculos. Si

no es holista —si los enunciados se corresponden *uno por uno* con los conjuntos de expectativas— afronta dificultades para explicar el carácter real del desarrollo de las teorías, el cual es decididamente 'holista'. Si, por otro lado, convierte a la *verdad* (o a la aceptabilidad) en un predicado de extensos sistemas de enunciados —tal vez cuerpos enteros de conocimiento— y no de oraciones aisladas, tendrá problemas para dar cuenta de la confiabilidad del aprendizaje (en verdad, me atrevería a afirmar que no puede explicarla, pero esto rebasa el objetivo de este ensayo).

Sin embargo, para el realista el enunciado (I) no es trivial. En su opinión, 'la nieve es blanca' no corresponde a un conjunto fijo de expectativas acerca del futuro, sino a que la nieve es blanca. En consecuencia, ¿cómo es posible que cuando creemos en determinados enunciados tendamos a actuar de tal forma que nuestras espectativas no se vean frustradas *en caso de que* esos enunciados resulten verdaderos (es decir, si se produjeran las situaciones correspondientes)?

La respuesta del realista es que la *relación* entre el surgimiento de la situación de que se trate y la satisfacción de nuestro objetivo es *en sí misma* algo sobre lo cual poseemos una gran cantidad de creencias verdaderas. Tenemos creencias verdaderas del tipo 'si hago X, obtendré Y'. ¿Cómo ocurre esto? Bueno, estas relaciones son *aprendidas*; por lo tanto, la explicación causal (programática) sobre la confiabilidad del aprendizaje *también* explica la existencia de creencias verdaderas acerca de las *relaciones* entre los actos y los objetivos en situaciones diversas.

Esta interpretación causal de la exactitud de los dos supuestos de la explicación mínima sobre la contribución de la conducta lingüística al éxito de la conducta total es lo que tenía en mente cuando al principio señalé que la *correspondencia* entre las palabras y las cosas, los enunciados y los estados de hecho constituye lo que explica el éxito del uso del lenguaje, aun cuando no se haga referencia a ella en el 'programa' para el empleo del idioma.

Hasta aquí dejé que pareciera un milagro el hecho de que la relación entre situaciones y enunciados descrita por la teoría causal de la percepción, la adquisición del lenguaje, etc., sea *además* la que está especificada por la definición de verdad para ese lenguaje. Pero también esto puede explicarse. Supongamos que C es una correspondencia arbitraria entre los enunciados y las situaciones, y llamemos VERDADERA (C) a una oración si se produce la situación con la

que sostiene la relación *C*. En términos generales, no existe entonces razón alguna para que la lógica deductiva preserve la propiedad de ser una VERDADERA (*C*). Pero parte de nuestra explicación sobre la confiabilidad de los hablantes consiste en que uno de los medios por los que adquirimos nuevas creencias es el uso de la lógica deductiva y que ésta preserva la *verdad*.[7] Para que tal explicación forme parte de la explicación total acerca de la confiabilidad, la correspondencia *C* debe pertenecer a una clase especial, y la elección más natural es la de una correspondencia que esté basada en una relación de satisfacción.

Desde luego, una vez seleccionada una correspondencia de esta clase para definir a la *verdad en L*, donde *L* es el lenguaje en cuestión, se hará posible definir la correspondencia de que se trate dentro de cualquier metalenguaje que posea la solidez suficiente y que contenga al propio L 'desentrecomillado', es decir, de tal manera que se satisfaga el criterio *T*. No obstante, observemos que (*T*) carece de significado en una traducción radical,[8] y en mis Conferencias John Locke argumenté que la definición de verdad para un lenguaje natural se halla subdeterminada por el tipo de consideraciones causales que discutimos aquí, además de que bien puede haber cierta relatividad del interés en nociones como 'verdad' y 'referencia' dentro del contexto de la traducción radical. Pero el hecho de que las explicaciones sean relativas con respecto al interés no implica que no puedan ser *correctas*, lo cual constituye lo único que importa para la argumentación que estoy desarrollando a favor del realismo.

¿La explicación causal es filosóficamente neutral?

Puedo imaginar que un filósofo contestaría a lo que acabo de exponer diciendo: 'Desde luego, usted ha delineado la explicación *científica* sobre la confiabilidad del aprendizaje', y la insinuación con-

[7] Si mi posición sobre la mecánica cuántica es correcta (véase *Mathematics, Matter and Method; Philosophical Papers*, vol. 1, Cambridge University Press, 1975, cap. 10), y si la lógica real del mundo no es la clásica sino la lógica modular descrita por Birkhoff y von Neumann, u otra parecida, tendríamos que decir, en lugar de lo expresado en el texto: 'la lógica deductiva clásica preserva la *verdad* en situaciones en que la mecánica cuántica puede ser ignorada'.

[8] Expongo esto, además de la relatividad del interés en las definiciones de verdad de la traducción radical en mis Conferencias John Locke (véase la Primera parte).

siste en que tal explicación *científica* debe ser compatible con *cualquier* postura metafísica. Ahora bien, es claro que todas las posturas metafísicas pretenden ser compatibles con todas las tendencias científicas, pero pretender algo no equivale a lograrlo.

Con el fin de entender cuáles son las dificultades afrontadas por el filósofo que no es realista y que desea incorporar la explicación causal sobre el éxito del uso del lenguaje a su imagen del mundo, discutiré una postura en particular. Se trata de la misma que describí en "What is 'Realism'?"[9] (una posición sugerida por algunas ideas de Michael Dummett). En ese artículo demostré que si dábamos un significado cuasi-intuicionista a las conectivas lógicas, preservaríamos la totalidad del discurso científico extensional y la semántica de Tarski (incluyendo al criterio *T*), al tiempo que evitaríamos cualquier compromiso con la existencia de entidades teóricas como los genes o las moléculas. De tal modo, a primera vista parecería que se puede asumir toda la explicación causal sobre la contribución de la conducta lingüística al éxito de la conducta total sin comprometerse realmente con la existencia de la mayoría de las entidades de las que habla la ciencia, o de una correspondencia entre las palabras y esas entidades. El realismo causal *en* la ciencia parecería ser compatible con una reinterpretación idealista de la ciencia total.

Pero como argumenté entonces, esto no es así. Pues la ciencia, según se la entiende habitualmente, realiza enunciados *modales*, enunciados sobre lo que es y lo que no es *posible*; por ejemplo, 'no se puede fabricar una máquina de movimiento perpetuo'. Y una reinterpretación de la ciencia, que no sea una mutilación, también debe 'traducir' estos enunciados (preservando el valor de verdad, así como las relaciones inductivas y deductivas). Pero sostengo que esto no puede hacerse.

Para comprender por qué no, comenzaré por señalar que la 'verdad', según el cuadro cuasi-intuicionista, se convierte en una versión de la aseverabilidad justificada. (Desde luego, es muy característico de las reconstrucciones idealistas el hecho de que los predicados ofrecidos como nociones de aceptabilidad para remplazar a la noción clásica de verdad resulten versiones de la aseverabilidad justificada.)

[9] "What is 'Realism'?", *Proceedings of the Aristotelian Society* (1975-1976), pp. 177–194.

Ahora bien, cuando se observa el color de una alfombra, la explicación causal debe contener un *margen de error*, el hecho de que sea *físicamente* posible que uno crea ver una alfombra verde, etc., y que la alfombra no sea verde. De esta forma:

(7) Puede justificarse la aserción 'la alfombra es verde', aun cuando la alfombra no sea verde,

constituye un enunciado modal implicado por nuestra teoría. Pero ello demuestra que la verdad *no puede ser* una aseverabilidad justificada.

Dummett calificó a mi argumentación como una especie de 'falacia naturalista' (se la podría denominar el 'argumento de la falacia idealista'), debido a su obvia similitud con la celebrada argumentación de Moore sobre la imposibilidad de definir el 'bien'. Lo que estoy afirmando es que para cualquier predicado *P* que el idealista quiera sustituir por 'verdadero', se puede hallar un enunciado *S* tal que de nuestra teoría causal del aprendizaje se siga lo siguiente:

(8) *S* puede tener la propiedad *P* y, sin embargo, no ser *verdadero*.

Y esto es así porque, sencillamente, la 'laguna' entre lo que es justificadamente aseverable, como quiera que esté construido, y lo que es *verdadero* (con el criterio *T* como único presupuesto) se encuentra 'incorporada' en nuestra teoría causal.

Sin embargo, parecería que Dummett tiene una fácil salida. Supongamos que el realista confirma los enunciados sobre lo que puede ser mediante la descripción de un 'modelo' en el cual se sostiene el enunciado en cuestión (porque cumple con las restricciones especificadas) y con la demostración de que el modelo satisface las *leyes* de nuestra teoría. Muy bien, supongamos que el filósofo que no es realista acepta lo anterior como procedimiento de verificación (el de aseverabilidad justificada) para los enunciados sobre lo que puede ser, e incopore estos últimos a *su* lenguaje con este mismo procedimiento de verificación.

Al extender así lo 'verdadero' (lo justificadamente aseverable) a los enunciados del tipo '*p* puede ser verdadero aun si *q*', quien no es partidario del realismo encontrará también que la proposición (7) es verdadera, y por lo tanto que:

(9) Puede ser aseverable justificadamente 'la alfombra es verde' aun cuando 'la alfombra es verde' no sea verdadera.

No resulta difícil mostrar cuál ha sido el error. Se reinterpretó la 'verdad' como aseverabilidad justificada (en cierto sentido). Y *aseverar que p* se reinterpretó como aseverar que *se justifica aseverar que p*. 'La alfombra es verde' se halla ahora *implicada por* 'se justifica la aserción de que la alfombra es verde'. De modo que si introducimos enunciados sobre lo que puede ser en el lenguaje según lo acabo de sugerir, estaremos dejando de lado la *incompatibilidad* entre *p implica q* y *p puede ser aun cuando no sea q*. No habremos preservado las relaciones deductivas entre los enunciados en el proceso de la traducción.

En realidad, el problema es incluso más grave: si se permitiera el uso de las nuevas implicaciones en la misma teoría científica, ésta se volvería incongruente. Asimismo, el término 'verdadero' no se intercambia con 'puede': '*p* puede ser aun si *q*' no equivale a '*p* puede ser verdadero aun si *q* es verdadero'.

He expuesto la posición de Dummett porque representa la más viable que conozco de las posturas ajenas al realismo. Los obstáculos que afronta revelan lo difícil que es proporcionar una reinterpretación idealista de nuestras teorías causales que sea capaz de preservar las relaciones deductivas e inductivas entre nuestros enunciados, incluyendo las oraciones *modales* que aceptamos, y de eludir, al mismo tiempo, mi argumento de la falacia idealista.

Nueva mirada al siglo XIX

La gran objeción del siglo XIX a la teoría de la verdad por correspondencia consistió en que no se podía pensar a la verdad como una correspondencia con los hechos (o la 'realidad') porque, según se decía, hacerlo así exigiría que uno fuera capaz de comparar directamente los conceptos con una realidad no conceptualizada, y los filósofos se sentían tan complacidos entonces como lo están ahora al señalar lo absurdo de semejante comparación.

Resulta innecesario indicar que ocurrían muchas cosas distintas con este argumento, demasiadas como para exponerlas aquí. (Por ejemplo, parte de su trasfondo era la opinión general de que no podemos percibir realmente los objetos materiales, opinión que ha sufrido fuertes ataques durante la segunda mitad de este siglo.) Pero algo que bien pudo haber figurado en las mentes de esos hombres es lo siguiente: supongamos, como es probable, que la comprensión de

un enunciado debe explicarse como *el conocimiento de sus condiciones de verdad*. Si la verdad es una correspondencia con la realidad, parecería que el conocimiento acerca de *cuál es la correspondencia* se halla presupuesto por el conocimiento de que tal y cual enunciado se encuentra o no se encuentra vinculado con algo en la relación en cuestión. Y si se hace equivaler la *comprensión del enunciado* al conocimiento de qué implica que ese enunciado esté o no vinculado con las entidades pertinentes en la relación en cuestión, el conocimiento acerca de cuál es la correspondencia se halla presupuesto en la comprensión de *todo* enunciado. ¿Pero en qué podría consistir este conocimiento, que no consiste en la aceptación de ningún *enunciado* porque precede a la comprensión de todos ellos?

La teoría semántica que subyace a esta formulación de la objeción, la de que la *verdad* es anterior al *significado* (es decir, que la verdad debe ser el concepto más importante para una teoría de la comprensión)[10] resulta bastante familiar en nuestros días. Lo que estoy proponiendo en el presente artículo es que rechacemos este punto de vista. Si concebimos a la comprensión del lenguaje como la posesión de una actividad racional en el uso del mismo, una actividad que implique 'reglas para las entradas del lenguaje' (procedimientos para supeditar algunas oraciones al control del estímulo),[11] protocolos para la inferencia deductiva e inductiva y 'reglas para las salidas del lenguaje' (procedimientos para, por ejemplo, 'tomar algo en nuestras manos cuando decidimos que lo óptimo sería tomarlo en nuestras manos'), el conocimiento implícito de las condiciones de verdad no estará *presupuesto* de ninguna manera por la comprensión del lenguaje. Para expresarlo sumariamente: no se necesita *saber* que existe una correspondencia entre las palabras y las entidades extralingüísticas para aprender nuestro lenguaje. Pero no por ello deja de existir esa correspondencia, la cual explica el *éxito* en lo que uno hace. *Después* de que uno ha aprendido su lenguaje, puede hablar

[10] Empleo aquí la terminología de las Conferencias William James (inéditas) de Michael Dummett. Aunque nuestras posiciones discrepan, el importante trabajo de Dummett me ha estimulado mucho, y este ensayo es en buena medida una respuesta al mismo.

[11] Nuevamente, la noción de 'control de estímulos' no es más que una excesiva simplificación o una idealización. Sin duda, el hecho de que la mente humana acepte una oración observacional depende, de una manera compleja, de la atención y de la teoría global, y no sólo de las rutinas inherentes o aprendidas para el reconocimiento de patrones.

sobre *cualquier cosa*, incluso de esa correspondencia. La opinión de Wittgenstein, expresada en su *Tractatus*, de que no puede describirse la correspondencia en cuestión, sino sólo 'demostrarla', es verdadera sólo en un sentido restringido: aun cuando se siga a Kripke, en lugar de a Tarski, de modo que se pueda hablar de la relación de referencia en el lenguaje objeto mismo y no sólo en un metalenguaje, todavía cabe definir una relación de referencia más amplia si se sube otro peldaño en el idioma.

Pero no obstante, aún es posible hablar de esta relación de referencia más amplia y no sólo 'demostrarla'. Y para los fines del tipo de explicación causal sobre la confiabilidad del aprendizaje, etc., que ya se discutió, la relación de referencia de la que podemos hablar en el lenguaje objeto mismo es más que suficiente. Como lo subrayó Carnap hace mucho tiempo, no hay nada 'indescriptible' en la relación que posee el lenguaje con el mundo.

Una objeción realista

Imagino que algunos realistas bien pueden sentirse incómodos con la explicación delineada en estas páginas acerca de la comprensión, incluso como modelo idealizado y excesivamente simplificado. ¿No es este modelo verificacionista? ¿Y no es el verificacionismo, en el fondo, una forma de idealismo?

Tomemos la primera acusación: en efecto afirmo que se ocultaba un importante concepto en el verificacionismo. El concepto consiste en que, en cualquier modelo verosímil de un hablante/oyente, el asignar a las oraciones una 'probabilidad' o una analogía no necesariamente cuantitativa de una 'probabilidad' va a desempeñar un papel crucial (aunque no es preciso asignarla a *cada* oración sobre una evidencia observacional, pues el lenguaje también se emplea para discutir qué *sucedería* en circunstancias cuya existencia quizás no podamos probar, por ejemplo, los fenómenos que se producen en el interior de los hoyos negros). En *Experience and Prediction*, Reichenbach ilustra el poder de lo que denomina la 'teoría probabilística del significado' de la siguiente manera: un fenomenalista tradicional tiene que explicar lo que se dice sobre objetos no observados —por ejemplo, árboles que no se ven— 'reduciéndolo' a lo que se dice sobre objetos observados (y en última instancia a las impresiones sensoriales) al estilo de Mach o Avenarius. Pero Reichenbach no

necesita proceder así. Una oración sobre árboles no observados (digamos, 'hay un árbol detrás de mí') se comprende si sabemos cómo asignarle un 'peso', es decir, cómo llegar a ella por inducción (Reichenbach no toma en cuenta la necesidad de una función de utilidad y una regla de decisión, las cuales se introducen en la obra posterior de Carnap sobre la probabilidad), cómo deducir/inducir consecuencias a partir de ella, etc. Por ejemplo, supongamos que veo la sombra de un árbol (como si perteneciera a un árbol que está atrás a mi izquierda). Antes he confirmado que: 'toda vez que se ve una sombra de árbol, hay un árbol en tal y cual relación espacial con esa sombra'. A partir de esta generalización confirmada por inducción y de la observación de la sombra de un árbol que está frente a mí, deduzco (o induzco, si la generalización es únicamente estadísitica) que: 'hay un árbol que guarda tal y cual relación espacial con *esta* sombra', y finalmente que 'hay un árbol detrás de mí'. Por lo tanto, ahora he aceptado (o asigné un peso a) una oración sobre un árbol no observado sin ninguna 'reducción' al estilo de Mach (o C. I. Lewis, o Russell, etc.) del árbol que no se ve a una serie de árboles observados, o impresiones sensoriales, reales o posibles. (Carnap presenta una explicación parecida en *Testability and Meaning*.)

En mi opinión, el punto correcto en lo anterior es que si no tuviéramos *ninguna* lógica inductiva, si sólo contáramos con el reconocimiento de los patrones y la lógica deductiva, careceríamos de bases para adscribirnos cualquier *concepto* sobre un 'objeto no observado'. Nuestra conducta lingüística se adecuaría a la explicación '"árbol" significa "árbol observado"', y en términos más generales, '"objeto" significa "objeto observado"'. En este sentido, nuestra lógica inductiva forma *parte de* nuestro concepto sobre un objeto no observado y, por lo tanto, de cualquier objeto. (Lo cual no significa que toda *modificación* de nuestra lógica inductiva sea una modificación de nuestro concepto de objeto; nociones como 'concepto' y 'significado' son una tosca cuadrícula sobrepuesta al uso.) De un modo semejante, nuestra lógica deductiva forma parte de nuestra comprensión acerca de qué es un conjunto o una propiedad, así como de lo que entendemos por cuantificadores y conectivas.

En segundo lugar, *no* se evidencia que *este* tipo de verificacionismo (la 'teoría probabilística del significado') sea incompatible con el realismo. (Creo que esta teoría es *incorrecta* por motivos completamente distintos, pero no está a discusión en este punto. El pro-

blema que por el momento me interesa examinar consiste en si dicha teoría es compatible o no con el realismo.)

Resulta evidente por qué la teoría del significado basada en la 'posibilidad de verificación concluyente en principio' es incompatible con el realismo. Éste supone que la verdad o la falsedad de nuestras oraciones depende (por lo general) de algo *externo*, es decir, de hechos que no son (lógicamente equivalentes a) hechos de la experiencia, sucesos que no están referidos a las sensaciones (o a las reglas del lenguaje, etc.). Pero el realismo, de ser verdadero, tiene que poderse expresar en uno u otro lenguaje ('si no lo puedes decir, tampoco lo podrás silbar.') Supongamos que *LR* es un lenguaje adecuado para expresar la tesis del realismo. Por ejemplo, quizás queramos dar en *LR* el tipo de explicación que he estado delineando acerca de que hay una correspondencia entre las oraciones y los hechos extralingüísticos (y no fenoménicos), cuya existencia explica la contribución que el uso del idioma aporta al éxito del comportamiento en general. Si el verificacionismo concluyente es correcto, debe haber condiciones de verdad fenoménicas para cada oración de todo lenguaje inteligible. (O bien, condiciones del 'lenguaje de las cosas observables', si el debate versa sobre el realismo en relación con las 'entidades teóricas', y no en relación con las entidades no fenoménicas. Pero pasaré por alto esta sutileza.) Por lo tanto, en términos más específicos, las oraciones de LR que pretenden describir las situaciones supuestamente no fenoménicas que corresponden, digamos, a las oraciones del lenguaje natural poseen en efecto condiciones de verdad fenoménicas; es decir, después de todo estas situaciones no son en realidad 'no fenoménicas'. En consecuencia, el realismo es falso o (peor) inexpresable.

Pero no se puede presentar un argumento así contra la compatibilidad entre el realismo y la 'teoría probabilística del significado'. ¿Por qué habrá de ser imposible desde el punto de vista realista que (1) toda oración significante posea uno u otro peso dentro de una situación observable, y (2) toda *diferencia* de significado se refleje en cierta diferencia de peso dentro de una situación observable? (De acuerdo con mi formulación, éstos son los dos principios de la 'teoría probabilística del significado' de Reichenbach.) Al menos, si ello es así o no, debería ser una pregunta abierta para el realista en cuanto realista, y no algo descartado por el realismo.

Sin embargo, algo muy importante que se vincula con lo anterior es que la 'teoría probabilística del significado' no exige que toda

oración significante posea un peso *elevado* (que 'se verifique') en una u otra situación; pues ello descartaría oraciones como 'hay una montaña de oro, pero "hay una montaña de oro" nunca tendrá un gran peso en las evidencias'. E insistir en una relación *semántica* entre la existencia y el conocimiento resulta inaceptable para el realista. De forma semejante, sería inadmisible para éste la exigencia de que toda oración significante sea 'falseable en principio' (que posea poco 'peso' en una situación observable). Pero ni Carnap ni Reichenbach imponen estos requisitos.

Sin embargo, es en efecto inaceptable para el realista tomar la 'teoría probabilística del significado' sencillamente por una *explicación* de lo significante, como lo hace Reichenbach. Supongamos que S_1 y S_2 son oraciones que describen dos estados de hecho que resultan posibles según la teoría física y no equivalentes según las leyes de la teoría física. Antes de que el realista pueda aceptar la 'teoría probabilística del significado', *una* de las cosas que habría que demostrarle es que todo par S_1, S_2 tiene la propiedad de que en una u otra situación observable S_1 y S_2 difieren en su peso. (Si admitimos que las diferencias de peso de la evidencia *no* observacional cuenten como diferencias de significado, este problema quizás desaparecería, pero entonces la teoría dejaría de ser verificacionista.) El realista no se hallaría dispuesto a estipular simplemente que S_1 y S_2 *cuentan* como la misma situación sólo *porque* adquieren el mismo peso en todas las evidencias observacionales.[12]

Aún más, como quiera que sea, la 'teoría probabilística del significado' me parece incorrecta como explicación del *significado*. En su formulación actual, no puede incorporar la división lingüística del trabajo (el hecho de que los procedimientos para confirmar *ser de oro, ser de aluminio, ser un olmo* o *ser David* no pertenecen a todo hablante, pues los hablantes difieren de los especialistas en cuanto a la fijación de la referencia en un enorme número de casos) ni la contribución del ambiente (el hecho de que en ocasiones la extensión de

[12] La presencia de una especie de incompletitud godeliana en cualquier lógica *inductiva* formalizada también representa una gran dificultad para la 'teoría probabilística del significado'. ¿Cuál de las posibles lógicas inductivas formalizadas debemos aplicar para calcular el 'peso'? (La incompletitud godeliana de la lógica inductiva fue señalada por primera vez en mi *Mathemathics, Matter, and Method; Philosophical Papers*, vol. 1, Cambridge University Press, 1975, cap. 17, y también se discute en el capítulo 18 del mismo libro.)

un término determina su significado y no a la inversa).[13] Desde mi punto de vista, los criterios utilizados por los especialistas para indicar si algo es o no es oro no forman 'parte del significado' de *oro* (es decir, la palabra no modifica su significado dentro del lenguaje si los especialistas adoptan un conjunto distinto de pruebas para identificar al mismo metal), aunque son parte de un mecanismo para determinar la extensión de ese término. Sin embargo, si no intervienen en el *significado* de *oro*, no existe nada en el significado que nos permita asignar *individualmente* un 'peso' a 'tal y cual cosa es oro', en oposición a la teoría de la que estamos hablando.

Pero no argumenté que deberíamos aceptar esta teoría, sino sólo que el modelo de Reichenbach y Carnap es parcialmente correcto como explicación del *uso* y la *comprensión*, no como explicación del *significado*.

Uso y significado

Si la interpretación que aquí he presentado es correcta, resulta evidente por qué deseamos una noción de la referencia, ¿pero por qué necesitamos una noción del significado?

Requerimos de una noción de la referencia debido a que el referente de un término es importante en muchas situaciones distintas. La compra y venta de vino representa un 'juego del lenguaje' completamente diferente a la discusión sobre vinos durante la sobremesa, pero la extensión del término 'vino' y las condiciones de verdad para 'se puede comprar vino de Israel en nuestro país' pueden ser las mismas en ambos contextos. Y *se debe a que* una oración puede tener las mismas condiciones de verdad en contextos distintos el que sea posible utilizar la misma lógica inductiva y deductiva en contextos diferentes. Pero, repito, ¿por qué necesitamos una noción de significado?

En mi opinión,[14] un lenguaje construido y empleado por un ser que no pertenece a ninguna comunidad no tendría necesidad de un concepto como el de *significado* de un término. Basta con determinar la referencia de cada término (simple y definido) y con describir lo que el usuario de ese lenguaje cree con respecto a cada palabra.

[13] *Cfr.* "The Meaning of 'Meaning' ".

[14] Para una exposición detallada, véase "The Meaning of 'Meaning' ".

Si el lenguaje cambia de tal forma que también se altera la referencia de un término, podemos decir que esto ha sucedido, o si el hablante revisa algunas de sus creencias podemos afirmar que esto ha ocurrido; ¿pero cuál es el objeto de señalar que algunas (aunque no todas) de estas últimas modificaciones cuentan como 'cambios de significado'?

Sin embargo, tan pronto como el lenguaje se convierte en un instrumento comunal, las cosas cambian. ¿Cómo podría tener lugar una *discusión* si no estamos en posibilidades de presuponer *nada* acerca de lo que creen los hablantes? ¿Puedo emplear con seguridad la palabra 'tigre' cuando hablo con alguien si, hasta donde sé, esta persona piensa que los tigres son una especie de *almejas*? ¿Adónde comenzaría la conversación?

En 'The Meaning of "Meaning" ' argumenté que el significado es un problema de varios componentes. Desarrollé el punto de vista de que *uno* de sus componentes es la *referencia* (extensión). (En mi opinión, el significado sólo fija a la referencia en tanto ésta es un *componente de* aquél, pero no por cuanto el significado sea un mecanismo para fijar a la referencia. Los verdaderos mecanismos para efectuar esto último, por ejemplo, los criterios utilizados por los especialistas para determinar si algo es o no es oro, *no* siempre forman parte del significado.) Otro componente importante, a mi entender, es el *estereotipo*, y los estereotipos no son más que conjuntos estandarizados de creencias o creencias idealizadas que se vinculan con los términos. (Por ejemplo, la creencia de que los tigres tienen normalmente rayas anaranjadas y negras forma parte de nuestro estereotipo de un tigre.) *No* necesitamos primordialmente a los estereotipos para fijar las extensiones; los especialistas pueden lograr esto y a menudo lo hacen mediante el empleo de criterios que no forman 'parte del significado' en modo alguno. El estereotipo vinculado con 'oro', por ejemplo, carece de toda utilidad para fijar la extensión de esa palabra (o en lo que a ello respecta, su extensión en los mundos posibles). *El lenguaje no sólo se utiliza para verificar, falsear y clasificar*, sino también para *discutir*. La existencia de estereotipos estandarizados, y por lo tanto del significado, es necesaria para la *discusión*, no para la clasificación.

Por cierto, Wittgenstein tenía razón cuando dijo que el lenguaje es una mezcolanza, *en el sentido de que* poseemos muchos patrones distintos para diversos tipos de discusión, y ello se refleja en el hecho de que la *cantidad* de información contenida en los 'significa-

dos', su *naturaleza* y su 'rigor' lógico (la aproximación a lo analítico) varía enormemente de una clase de palabra a otra e incluso entre las palabras de la misma clase. (Comparemos 'soltero' con 'oro', y también 'oro' con 'molibdeno'. Estos últimos son nombres de metales, pero existe un rico estereotipo para el oro y casi ninguno para el molibdeno, fuera de la característica de ser 'metal'. Nuevamente, 'tigre' y 'comadreja' son nombres de animales, pero el estereotipo —y en mi opinión, el significado— de 'tigre' incluye la característica de que tiene rayas anaranjadas y negras, mientras que no sé de qué color son las comadrejas y, aún más, no lo sé en virtud de 'conocer el significado' de 'comadreja'.)

Moralejas

Si cabe sacar algunas moralejas de mi exposición, éstas quizás sean: (1) que la noción de que se aprende el lenguaje natal mediante el aprendizaje de *cuáles son las condiciones de verdad* para sus diversas oraciones carece por el momento de un sentido inteligible, al menos para el realista; (2) que no se concluye que las nociones sobre la verdad y la referencia del realista carecen de importancia para la discusión acerca del lenguaje, sino que tienen importancia para la explicación de la contribución que la conducta lingüística aporta al éxito de la conducta total, no para una teoría de la comprensión; y (3) puesto que (según lo expuse en otro lugar) la noción de 'significado' no posee ni la naturaleza ni la función que los filósofos suponían, parece más probable que la introducción de la palabra 'significado' a las discusiones sobre la comprensión y el uso confundan las cosas en lugar de aclararlas.

CUARTA PARTE

REALISMO Y RAZÓN

REALISMO Y RAZÓN*

Según una forma de concebirlo, el realismo es una teoría empírica.[1] Uno de los hechos explicados por ésta es que las teorías científicas tienden a 'converger', en el sentido de que las primeras teorías son, con mucha frecuencia, casos limitantes de las posteriores (y ello constituye la razón de que podamos considerar que los términos teóricos preservan sus referencias a través de la mayor parte de los cambios de las teorías). Otro de los hechos que explica consiste en la cuestión más mundana de que el uso del lenguaje contribuye a la obtención de nuestros objetivos, de la satisfacción o de lo que se desee.

En pocas palabras, la explicación realista no consiste en que el lenguaje refleja al mundo sino en que los *hablantes* reflejan al mundo, es decir, a su ambiente, en el sentido de que *construyen una representación simbólica de ese ambiente*. En 'Referencia y comprensión'[2] señalé que la 'correspondencia' entre las palabras y los conjuntos de cosas (formalmente una *relación de satisfacción*, de acuerdo con Tarski) puede ser considerada como parte de un *modelo explicativo* de la conducta colectiva de los hablantes.

No voy a revisar esto aquí, pero me tomaré la libertad de referirme a este tipo de realismo —la aceptación de esta clase de imagen científica sobre la relación que guardan los hablantes con su medio y sobre el papel del lenguaje— como realismo *interno*.

* Discurso del presidente a la División del Este de la Asociación Filosófica Estadounidense, Boston, Mass., 29 de diciembre de 1976.

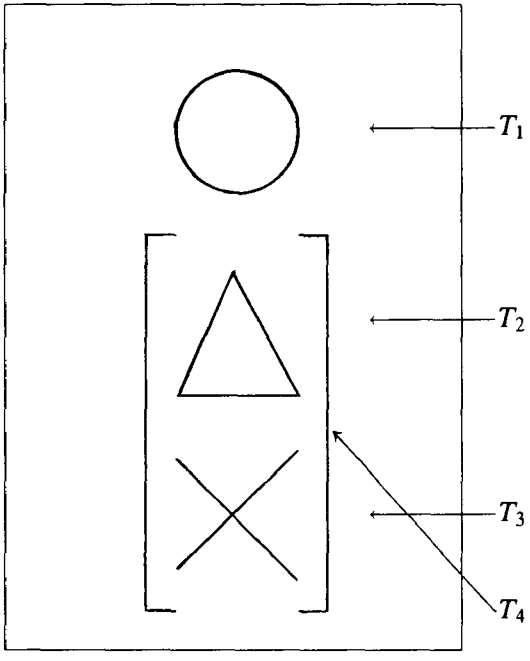
[1] Esto se explica en mi "What is 'Realism'?", en *Proceedings of the Aristotelian Society* (1975–1976), pp. 177–194. Primera parte, segunda conferencia de este volumen.

[2] Dictada en Jerusalén en mayo de 1976; tercera parte de este volumen.

El realismo *metafísico*, por otro lado, es menos una teoría empírica que un modelo, y un 'modelo' parecido a 'bolas de billar en colisión'. Es, o pretende ser, un modelo de la relación que posee *cualquier* teoría correcta con todo EL MUNDO o con una parte de él. He llegado a la conclusión de que se trata de un modelo incoherente, y esto es lo que quiero compartir con ustedes.

Tracemos la forma básica de este modelo.

EL MUNDO



En su forma primitiva, hay una relación entre cada término del lenguaje y una porción de EL MUNDO (o una *clase* de porción, si se trata de un término general).

Esta relación, la de referencia, es suministrada *por la verdad*, la *semántica condicional* del lenguaje en las versiones canónicas de la teoría; es decir, *comprender* un término, digamos T_1, consiste en saber a qué porción de EL MUNDO se refiere (o, en algunas versiones, conocer una condición necesaria y suficiente para que se refiera a una porción de EL MUNDO). No voy a presuponer en lo que sigue

que esta explicación de la comprensión forma parte del cuadro, aunque ciertamente fue el presupuesto de los realistas metafísicos en el pasado

Sin embargo, tiene que *existir* por lo menos una relación de *referencia* determinada entre los términos de *L* y las porciones (o conjunto de porciones) de EL MUNDO en el modelo realista metafísico, ya se suponga que *comprender L* consiste en 'conocer' esa relación o no. Lo que diferencia a este cuadro del realismo *interno* (que utiliza un cuadro parecido *dentro* de una teoría) es que (1) se supone que se aplica a *todas* las teorías correctas al mismo tiempo (de modo que sólo puede ser desarrollado con una 'ambigüedad característica', es decir, trasciende la formalización completa de una teoría en particular), y (2) se supone que EL MUNDO es *independiente* de cualquier representación personal que tengamos de él; en realidad, se sostiene que quizás seamos *incapaces* de representarnos correctamente a EL MUNDO (por ejemplo, el realista metafísico nos dice que tal vez seamos 'cerebros en una tina').

La consecuencia más importante del realismo metafísico es que supone que la *verdad* es *radicalmente ajena a lo epistémico*; podríamos ser 'cerebros en una tina' y, por lo tanto, *sería falsa* la teoría que se considera 'ideal' desde el punto de vista de la utilidad operativa, la belleza y la elegancia internas, la 'verosimilitud', la simplicidad, el 'conservadurismo', etc. Lo 'verificado' (en cualquiera de sus sentidos operativos) no implica a lo 'verdadero', en el cuadro realista metafísico, ni siquiera en el límite ideal.

Esta característica es la que distingue al realismo metafísico, según estoy empleando el término, de la simple creencia de que *existe* una teoría ideal (el realismo de Peirce) o, con un fundamento más débil, que la teoría ideal constituye un ideal regulador, presupuesto por las nociones de 'verdadero' y 'objetivo', de acuerdo con la acepción clásica de estos vocablos. Y es justamente la característica que pretendo atacar.

De modo que supongamos que T_1 es, desde nuestra perspectiva, una teoría ideal. Si eliminamos las restricciones de nuestros poderes reales, demasiado finitos, podemos imaginar que T_1 posee todas las propiedades que queramos *salvo la de verdad objetiva*, la cual queda abierta. Por ejemplo, podemos pensar que T_1 está completa, es coherente, que (hasta donde sabemos) predice correctamente todas las oraciones observacionales, que satisface cualquier 'limitación operativa' existente (si son 'confusas', supongamos que T_1 las satisface

con claridad), que es 'hermosa', 'simple', 'verosímil', etc. El supuesto bajo consideración consiste en que T_1 puede ser todo lo anterior *y no obstante ser* (en realidad) *falsa*.

Doy por sentado que EL MUNDO tiene (o puede descomponerse en) infinidad de porciones. También que T_1 *dice* que hay una infinidad de cosas (así, en *este* aspecto *es* 'objetivamente correcta' acerca de EL MUNDO). Ahora bien, de acuerdo con la hipótesis, T_1 es *coherente* y tiene (solamente) modelos infinitos. En consecuencia, por el teorema de la completitud (en cuanto modelo teórico), T_1 posee un modelo para cada cardinalidad infinita. Tomemos un modelo M que esté en la misma cardinalidad de EL MUNDO.[3] Proyectemos, uno por uno, a los individuos de M sobre las porciones de EL MUNDO y utilicemos este mapa para definir las relaciones directas de M con EL MUNDO. El resultado es una relación de satisfacción SAT, una 'correspondencia' entre los términos de L y los conjuntos de porciones de EL MUNDO, de tal modo que la teoría T_1 se vuelve *verdadera* con respecto a EL MUNDO, siempre que interpretemos 'verdadero' como VERDADERA(SAT).[4] De esta forma, ¿en qué queda la afirmación de que incluso la teoría *ideal* T_1 puede ser *realmente* falsa?

Bueno, se podría argumentar que SAT no es la correspondencia *prevista* entre L y EL MUNDO. ¿Qué significa 'prevista' en este contexto?

T_1 tiene la propiedad de satisfacer todas las restricciones *operativas*. De manera que si 'hay una vaca frente a mí en tal y cual momento' pertenece a T_1, la oración *parecerá* sin duda verdadera; será 'exactamente como si' hubiera una vaca frente a mí en ese momento. Pero SAT es una interpretación *verdadera* de T_1. T_1 es VERDADERA(SAT), de modo que 'hay una vaca frente a mí en tal y cual momento' es 'verdadera' en el sentido de ser VERDADERA(SAT).

Por otro lado, si 'hay una vaca frente a mí en tal y cual momento' es *operativamente* 'falsa' (falseada), la oración es una FALSA(SAT).

[3] Si EL MUNDO es finito, supongamos que la teoría es compatible con la existencia de únicamente N individuos (donde N es la cardinalidad de EL MUNDO) y escojamos un modelo con N individuos en lugar de utilizar el teorema más estricto sobre el modelo teórico al que se alude en el texto.

[4] Si en este caso SAT es una relación del mismo tipo lógico que 'satisface', se supone que VERDADERO(SAT) estará definido en términos de SAT exactamente como 'verdadero' está definido (por Tarski) en términos de 'satisface'. Así, 'VERDADERO(SAT)' es la propiedad de verdad 'determinada' por la relación SAT.

Por lo tanto, la interpretación de la 'referencia' en *L* como SAT satisface sin duda todas las restricciones *operativas* de la referencia. Pero la interpretación de la 'referencia' como SAT cumple también con todas las restricciones *teóricas* de la referencia. Hace que la teoría *ideal* T_1 resulte *verdadera*.

En consecuencia, ¿cuáles son las *otras* restricciones de la referencia que podrían identificar a otra interpretación como (la única) 'prevista' y a SAT como una interpretación 'no prevista' (tomando el término 'interpretación' en el sentido del modelo teórico)? El supuesto de que incluso una teoría 'ideal' (desde el punto de vista pragmático) podría ser *realmente* falsa parece derrumbarse en la *ininteligibilidad.*

Observemos que una teoría 'causal' de la referencia no es de ninguna ayuda en este punto (no podría serlo), pues de qué manera sería posible que las 'causas' tuvieran una referencia exclusiva resulta tan desconcertante como preguntarnos de qué manera podría tenerla el término 'gato' en una representación metafísica realista.

En cierta medida se puede rastrear el problema hasta Occam, quien introdujo la idea de que los conceptos son *particulares* (mentales). Si son particulares ('signos'), cualquier concepto que obtengamos de la *relación* entre un signo y su objeto es *otro signo*. Pero desde mi punto de vista, resulta ininteligible que el tipo de relación que, en la concepción del realista metafísico, se sostiene entre un signo y su objeto pueda distinguirse ya sea exponiendo el signo, así:

VACA

o exponiendo otro signo distinto, así:

REFIERE

o tal vez

CAUSA

Por otro lado, si los conceptos no son particulares, la posibilidad obvia es que (en tanto los tenemos 'en la cabeza') representen *modos de usar* signos. Pero una teoría sobre el 'uso', aunque inteligible (y creo que correcta) como explicación acerca de en qué consiste

comprender los signos, *no identifica una relación exclusiva* entre los términos de T_1 y los 'objetos reales'. Si no creemos que los conceptos sean, o bien particulares (signos), o bien modos de utilizar a los signos, me parece que vamos a regresar a la captación directa (y misteriosa) de las formas.

Supongamos que nosotros (y todos los otros seres sensibles) somos y siempre fuimos 'cerebros en una tina'. ¿Cómo ha llegado a suceder que *nuestra* palabra 'tina' se refiera a tinas *nuoménicas* y no fenoménicas.

Si lo anterior no es sólo una nueva antinomia, se debe demostrar que hay por lo menos una postura inteligible en la cual *no* se presenta. Y esa postura existe. La antinomia no surge en la postura que Michael Dummett ha estado defendiendo. Intentaré explicarme:

La idea de Dummett[5] consiste en elaborar la *teoría de la comprensión* de acuerdo con las nociones de *verificación y falsificación*. A esto denomina 'semántica no realista'.

La diferencia entre esto y el viejo fenomenalismo estriba en que no hay ninguna 'base' de hechos positivos (por ejemplo, datos de los sentidos) con respecto a la cual uno emplee en última instancia la semántica de las condiciones de verdad, la lógica clásica y las nociones *realistas* de verdad y falsedad. La analogía se establece con el intuicionismo matemático: el intuicionista utiliza *su* noción de 'verdad' —la demostrabilidad constructiva— *aun cuando esté hablando sobre la prueba constructiva misma*. Comprender una oración, en esta semántica, equivale a saber en qué consiste su prueba (verificación). Y ello se aplica *incluso a los enunciados que describen a las verificaciones*. De tal manera, puedo tomar como oración primitiva 'tengo un dato sensorial rojo' o 'veo una vaca' o, si empleo la semántica desde el punto de vista del cerebro en vez del de la persona, 'tales y cuales neuronas funcionaron'.

Cualquiera que sea el lenguaje que yo use, la oración primitiva, por ejemplo, 'veo una vaca', será aseverable si y sólo si está *verificada*. Y decimos que está verificada *al emitir la oración* 'veo una vaca'. Para

[5] Esto se encuentra completamente detallado en sus Conferencias William James (inéditas). Una explicación parcial aparece en su contribución a la conferencia sobre 'Language, Intentionality and Translation Talk', reimpresa en *Synthese*, vol. 27, núms. 3/4, julio-agosto de 1974.

utilizar un término de Roderick Firth, 'veo una vaca' es una 'autojustificación' en este tipo de epistemología, no porque resulte *incorregible*, ni siquiera necesariamente en el sentido de hallarse determinada por completo (es decir, de que obedece a una estricta bivalencia: ser verdadera o falsa categóricamente). (Según Dummett, los hechos son 'dúctiles hasta el fin' en este cuadro.) Lo importante es que los conceptos realistas de verdad y falsedad no poseen ninguna utilidad en esta semántica.

Ahora bien, no se presenta aquí el rompecabezas acerca de qué distingue a una correspondencia como *la* relación de referencia. La noción de 'referencia' no se emplea en esta semántica. Podemos introducir 'se refiere' en el lenguaje al estilo de Tarski, pero entonces:

(1) 'Vaca' se refiere a las vacas,

lo cual no sería más que una tautología, y la *comprensión* de (I) no hace ninguna referencia a la representación metafísica realista.

Un punto importante. No tiene caso aplicar la semántica no realista (prefiero llamarla *semántica verificacionista*, porque no es incompatible con el realismo *interno*) de acuerdo con un nivel de 'hechos positivos', incluso de datos sensoriales. Pues si se tratara a estos últimos como 'datos positivos', es decir, si la semántica verificacionista es presentada en un metalenguaje *para el cual* damos la explicación *basada en las condiciones de verdad* sobre la comprensión, podemos repetir toda la objeción a la inteligibilidad del realismo metafísico (como una objeción a la inteligibilidad del *meta*lenguaje); pensemos sencillamente en los datos sensoriales *pasados* (o *futuros*) como la parte 'externa' de EL MUNDO. (Lo anterior es una reconstrucción de uno de los aspectos del argumento sobre el lenguaje privado de Wittgenstein.) Ésta es la razón de que el movimiento de Dummett dependa del uso de la semántica verificacionista en todo su camino ascendente (o descendente), en el metalenguaje, el metametalenguaje, etc.

El motivo por el que me interné en este problema es el siguiente: en 'Referencia y comprensión' argumenté que se puede proporcionar el *modelo de un hablante* del lenguaje de acuerdo con la noción de 'grado de confirmación' (que podría recibir el nombre más apropiado de 'grado de verificación' cuando desempeña esta función de comprensión teórica). Y afirmé que las nociones realistas de la verdad y la referencia no intervienen para explicar lo que ocurre 'en

las mentes' de los hablantes, sino para *explicar el éxito* del uso del lenguaje. Por lo tanto, instaba a que aceptáramos una especie de semántica 'verificacionista'. (Aunque no en el sentido de la teoría verificacionista del *significado*, pues, como lo expuse en otro lugar,[6] el 'significado' no es simplemente una función de lo que sucede 'en nuestras mentes', sino también de la *referencia*, y ésta se encuentra determinada por prácticas *sociales* y por paradigmas físicos reales, no sólo por lo que acontece en el interior de un hablante en particular.) Pero además afirmaba que se puede seguir siendo un *realista* aun cuando se acepte este modelo 'verificacionista'. Porque la aseveración del realista de que existe una correspondencia entre las palabras y las cosas no es *incompatible* con una explicación 'verificacionista' o 'utilitaria' sobre la comprensión. Tal correspondencia, en mi opinión, forma parte de una *teoría explicativa* de la interacción de los hablantes con su ambiente.

Lo cierto es que Dummett y yo *estamos de acuerdo* en que no se puede considerar la comprensión de una oración (en general) como el conocimiento de sus condiciones de verdad, porque entonces se vuelve ininteligible en qué consiste, a su vez, *ese* conocimiento. Ambos *coincidimos* en que la teoría de la comprensión debe elaborarse en términos verificacionistas. (Aunque no creo que esta teoría pertenezca por entero a una teoría del *significado* que no sirve *en este caso*; la teoría del significado, en mi opinión, presupone a la teoría de la comprensión *y de la referencia*, y en la referencia estriba todo el problema.) Pero ahora parece que al conceder que debe presentarse *alguna* clase de semántica verificacionista como explicación de la comprensión (o de la 'competencia lingüística' en el sentido de Chomsky), he proporcionado a Dummett las armas que necesitaba para demoler al realismo metafísico, ¡la imagen a la que yo me aferraba!

¿Y qué? En este punto, *mi* opinión consiste en que la respuesta natural sería la siguiente: el realismo metafísico, por lo tanto, se derrumba. Pero no así (afirmo) el realismo interno —la teoría empírica de 'Referencia y Comprensión'. De todos modos, el primero no era más que una *representación*. Si ésta resultaba en realidad incoherente, sin duda la moraleja *no* es que hay algo mal en el realismo *per*

[6] *Cfr.* mi "The Meaning of 'Meaning'", en mi *Mind, Language and Reality; Philosophical Papers*, vol. 2, Cambridge University Press, 1975.

se, sino simplemente que el realismo *equivale* al realismo interno. *Y el realismo interno es todo lo que queremos o necesitamos.*

En verdad, creo que esto es enteramente cierto. Pero la moraleja no acaba aquí. El realismo metafísico se derrumba *en un punto particular* (voy a argumentar que también se viene abajo en otros lugares). Y el punto en que se derrumba nos dice algo. El realismo metafísico se cae justo cuando afirma que se distingue del realismo de Peirce, es decir, de la pretensión de que existe una teoría ideal (no estoy diciendo que incluso *esa* afirmación no deje de resultar problemática, pero lo es de una manera diferente). Dado que Peirce mismo y los verificacionistas siempre *sostuvieron* que el realismo metafísico se vuelve incoherente *justo* en ese punto, y que realistas como yo pensábamos que estaban *equivocados*, no hay modo de evitar la desagradable admisión de que 'ellos tenían razón y nosotros nos equivocamos', por lo menos en lo que respecta a una cuestión sustancial.

Quiero referirme ahora a otros puntos en los cuales el cuadro del realista metafísico resulta incoherente. Consideremos el universo simple que sigue: supongamos que 'el mundo' es una *línea recta*, así:

(Si se quiere, puede haber en la línea personas unidimensionales, con mis disculpas a Marcuse. Cómo diferenciar a los niños de las niñas es algo que ignoro.)

Consideremos estas dos historias acerca de EL MUNDO:[7]

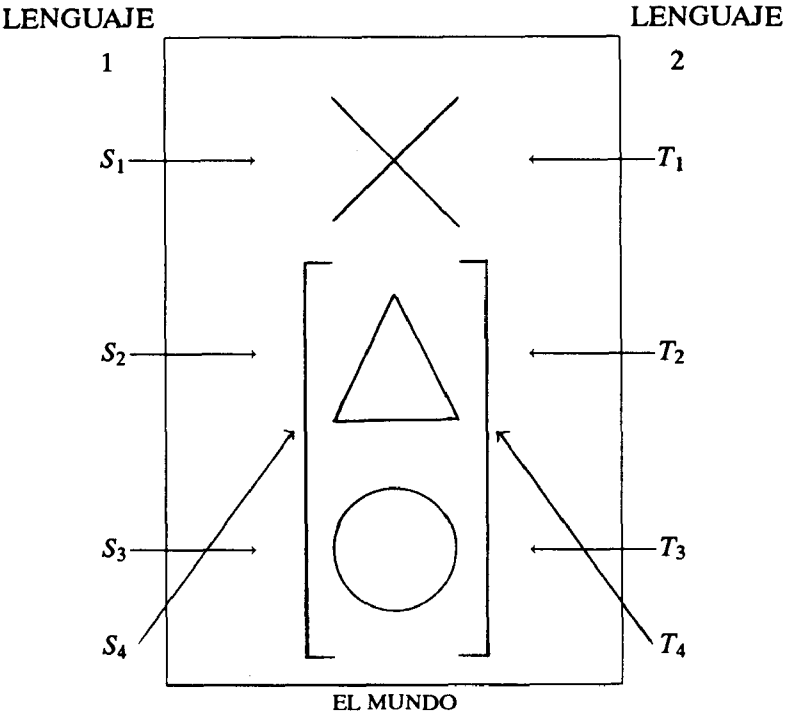
Historia 1. *Tiene puntos*, es decir, la línea posee partes que son *segmentos* y también partes infinitamente pequeñas llamadas 'puntos'. Se sostiene la *misma relación* —'ser parte de'— entre los puntos y los segmentos lineales que los contienen, y entre estos segmentos y los segmentos mayores de la línea (y entre cualquier porción y la totalidad de la línea).

Historia 2. *No tiene puntos*; la línea y sus partes poseen siempre *extensión*. 'Por supuesto', dice el narrador de esta historia, 'no estoy sugiriendo que la historia 1 sea *falsa*. Sólo hay que entender que los *puntos* son construcciones lógicas de los segmentos de la línea. Lo

[7] Compárense los sistemas alternativos discutidos en Nelson Goodman, *The Structure of Appearance*, Reidel, Dordrecht, 1951, 1977, cap. 1, secciones 1–4.

que se dice sobre los puntos deriva en gran medida de lo expresado sobre los conjuntos de segmentos convergentes.'

Un realista 'duro' quizás afirme que existe un 'hecho positivo' con respecto al cual la historia 1 o la historia 2 es verdadera. Pero los 'realistas sofisticados', como los he denominado, conceden que ambas son 'descripciones equivalentes'. En efecto, esto admite que los segmentos de la línea constituyan un conjunto adecuado de 'invariantes': una descripción de EL MUNDO que explique lo que ocurre en cada segmento es una descripción *completa*. Anteriormente, mi opinión era que esto no representaba un problema para el realista y afirmaba que se equiparaba al hecho de que pueden hacerse mapas de la tierra a partir de 'proyecciones' diferentes (Mercator, Polar, etc.). El cuadro realista metafísico queda ahora como el que sigue:



En particular, creo que puede darse el caso de que sobre el mismo objeto real —aunque, desde luego, no dentro de la misma teoría—

se elabore el mapa de lo que representamos como términos 'incompatibles'. Así, el objeto real que se denomina 'punto' en una teoría, tendría que ser llamado 'conjunto de segmentos convergentes de la línea' en otra. Y podría proyectarse el *mismo* término sobre un objeto real en una teoría y sobre uno diferente en la otra. Sostengo que es propiedad del mundo mismo, a saber, propiedad de EL MUNDO mismo, el 'admitir estas proyecciones distintas'.

Como ha insistido Nelson Goodman durante muchísimos años, el problema radica en si es posible que esta explicación conserve a EL MUNDO, pero a costa de abandonar toda noción inteligible sobre *cómo* es. Cualquier oración que cambie los valores de verdad al pasar de una teoría correcta a otra, también correcta —por ejemplo, una descripción equivalente— sólo expresará una propiedad de EL MUNDO relativa a la teoría. Y en tanto haya más oraciones de ese tipo, serán más las propiedades de EL MUNDO que se volverán relativas a la teoría.

Por ejemplo, si aceptamos que la historia 1 y la historia 2 son descripciones equivalentes, la propiedad de *ser un objeto* (en oposición a ser una clase o un conjunto de cosas) se volverá relativa a la teoría. Consideremos ahora una tercera historia:

Historia 3. *Sólo los segmentos de la línea tienen puntos finales racionales* (es decir, como en esta historia no hay 'puntos', excepto en cuanto a construcciones lógicas, (1) cada segmento de la línea presenta una longitud racional; (2) la porción de la línea[8] que se encuentra entre dos segmentos cualesquiera es un segmento de línea y, por lo tanto, tiene una longitud racional; (3) cada segmento de la línea puede dividirse en *n* porciones iguales para cada *n* entero; (4) existe por lo menos un segmento de la línea; y (5) la unión de dos segmentos es un segmento de la línea). Los segmentos irracionales se tratan como construcciones lógicas, es decir, conjuntos de 'puntos', donde los 'puntos' son en sí mismos conjuntos convergentes de Cauchy de segmentos *racionales*.

Un realista 'duro' podría volver a presentar objeciones, esta vez debido a que la historia convierte a un segmento irracional de la línea en un tipo lógico diferente al segmento racional. Pero el defensor puede replicar: ¿Acaso no es algo común en matemáticas identificar a los objetos con conjuntos de otros objetos que son, desde el

[8] El lector matemático advertirá que en la historia 3 no hay distinción entre los segmentos de la línea *abiertos* y *cerrados*, porque no existen los *puntos*.

punto de vista *preanalítico*, del mismo tipo lógico? De tal forma, los enteros negativos y los positivos, los números enteros y los racionales, los racionales y los reales, los reales y los imaginarios son todos 'números' *en términos preanalíticos*, pero *cuando formalizamos* las matemáticas estamos acostumbrados a tratar los números negativos (o desde una perspectiva más general, los 'números con signo') como *pares ordenados* de 'números naturales', los *racionales* como *pares ordenados* de 'números con signo', los irracionales como *conjuntos* de racionales, etc. En consecuencia, ¿por qué no vamos a tratar los segmentos irracionales de la línea como conjuntos de conjuntos de segmentos racionales? Después de todo, los segmentos racionales de la línea son una base para la topología; si se sabe lo que está ocurriendo en cada segmento racional, se obtendrá una descripción completa de todos los sucesos, etc.

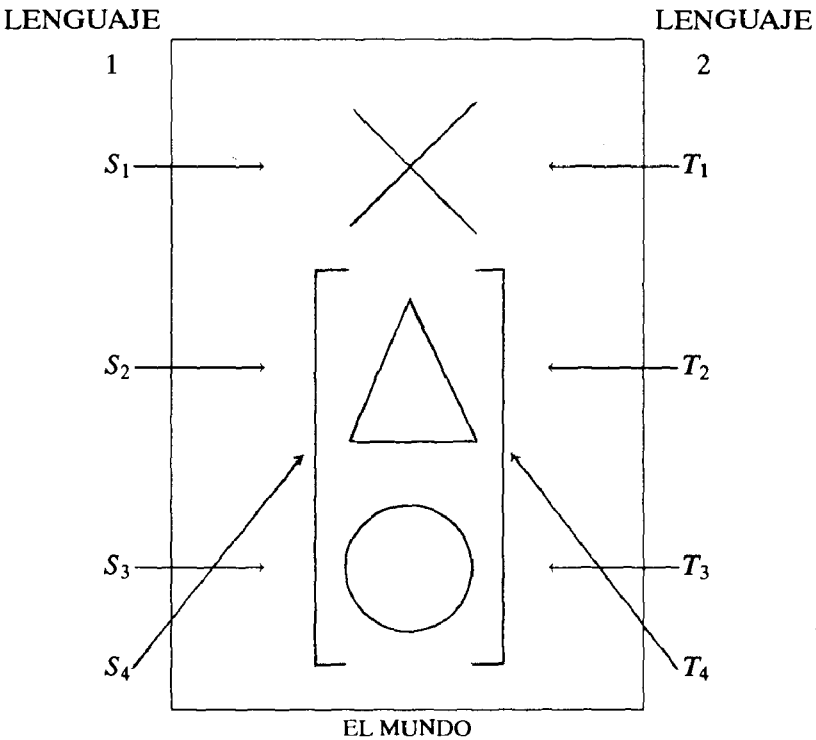
Pero si aceptamos la historia 3 como otra descripción equivalente más sobre 'el mundo', ¡incluso la *cardinalidad* de este último se vuelve relativa a la teoría, pues tenemos una cantidad de objetos sólo enumerable en la historia 3 y no enumerable en las historias 1 y 2. (Podemos tratar de evitar esto considerando también a los *conjuntos* como 'objetos' pero, como lo demostré en otro lugar, lo que se dice sobre los 'conjuntos' puede traducirse a expresiones sobre *posibilidades*.)

Lo anterior no es un artificio de mi sencillo ejemplo: la teoría física real se encuentra llena de instancias similares. Se pueden construir, por ejemplo, puntos espaciotemporales como objetos o como propiedades. También se pueden construir campos como objetos o hacer cualquier cosa con las partículas que actúan a distancia (utilizando potenciales retardados). El hecho es que *tantas* propiedades de EL MUNDO —comenzando por las que son *sólo categoriales*, como la cardinalidad, los particulares o los universales, etc.— se vuelven 'relativas a la teoría' que EL MUNDO acaba por parecerse a un mundo 'nuoménico' kantiano, una *simple* 'cosa en sí'. Si no se puede decir *cómo* es EL MUNDO independientemente de la teoría, las expresiones sobre todas estas teorías en tanto descripciones de 'el mundo' estarán vacías.

Otro punto en el cual el realista metafísico encuentra problemas

Esto se vincula con lo que Quine denomina 'relatividad ontológica'. Supongamos que concentramos nuestra atención, por el momento,

en las teorías *completas*. Si T es una teoría completa, podemos definir una relación de equivalencia para sus términos —*coextensión demostrable*—, con la propiedad de que si dos términos pertenecen a diferentes clases de equivalencias, no habrá *ningún* modelo de la teoría en que se refieran al mismo referente, mientras que si pertenecen a la misma clase de equivalencias, tendrán el mismo referente en *todos* los modelos de la teoría. Por lo tanto, para nuestros propósitos, podemos considerar que los términos son los mismos si se encuentran en la misma clase de equivalencias, es decir, si son 'coextensivos, tomando a la teoría según su valor nominal'. Hecha esta identificación preliminar, observamos que nuestro cuadro es correcto, y lo repito:



Así, sólo hay una *única* 'traducción' capaz de preservar a la referencia que está vinculando a los lenguajes.

Pero es obvio que existen con frecuencia interpretaciones relativas de una teoría con respecto a otra que *no son equivalentes*. Se puede interpretar la Historia 1 con respecto a la Historia 2 (en nues-

tro caso) de *muchas* maneras distintas. Los 'puntos' pueden ser, por ejemplo, conjuntos de segmentos de línea cuyas longitudes son potencias negativas de 2 o potencias negativas de 3.

Si el cuadro que representé fuera correcto, tendría que haber un 'hecho positivo' respecto a *qué* traducción preserva *realmente* a la referencia en cada uno de esos casos.

Así como, para responder a la objeción anterior, permitimos que el cuadro se complicara al admitir que el mismo término fuera proyectado sobre objetos reales distintos cuando aparece en teorías diferentes, también podemos *volver* a complicarlo para contestar a la segunda objeción: podemos decir que el lenguaje posee *más de una* forma correcta de ser proyectado en EL MUNDO (debe poseerla, por cuanto tiene más de una forma de ser proyectado correctamente sobre un lenguaje que está bien proyectado sobre el mundo). Pero ahora *toda* comprensión del cuadro parece desvanecerse: si lo que constituye un conjunto *único* de cosas *en una teoría correcta* no puede serlo 'en la realidad', la esencia del cuadro se derrumba.

¿Por qué lo anterior no es una refutación del realismo interno?

Supongamos que tratamos de desafiar al realista *interno* con la siguiente pregunta: '¿Cómo sabe que "vaca" se refiere a las vacas?' 'Después de todo,' señalamos, 'existen otras interpretaciones del lenguaje total de usted, innumerables interpretaciones (en el sentido de las relaciones de satisfacción), que convertirían en verdadera a una teoría *ideal* (en el lenguaje de usted). De hecho, supongamos que Dios nos dio el *conjunto de todas las oraciones verdaderas* en el lenguaje de usted (para este fin, pensemos que poseemos memorias infinitas). Llamemos a este conjunto la teoría *perfecta*. Aun *así* habría una cantidad infinita de interpretaciones admisibles de la teoría *perfecta*, interpretaciones que, como vimos, satisfarían *todas* las restricciones operativas y teóricas. Incluso la *oración* " 'vaca' se refiere a las vacas" es verdadera en todas estas interpretaciones. De modo que, ¿cómo sabe que es verdadera en el sentido de ser verdadera según la *única* interpretación 'prevista'? ¿Cómo sabe que "vaca" se refiere a las vacas en el sentido de que se refiere a *un solo* conjunto determinado de cosas, y no a un conjunto determinado de cosas *dentro de cada interpretación admisible*?' (Desde luego, esto equivale a objetar al realista interno exactamente de la misma forma en que objetamos al realista metafísico.)

El realista interno podría replicar que '"vaca" se refiere a las vacas' se concluye inmediatamente de la definición de 'se refiere'. De hecho, esa oración sería verdadera aun cuando el realismo interno fuera falso, aunque podríamos revisar el enunciado '"vaca" se refiere a las vacas' descartando la teoría misma (o por lo menos, descartando o impugnando la noción de una *vaca*), y de este modo el hecho de que '"vaca" se refiere a las vacas' no sea *absolutamente* irrevisable se manifiesta como *relativo a la teoría*, '"vaca" se refiere a las vacas' es una verdad lógica.

El crítico replicaría entonces que su pregunta no ha sido contestada. '"Vaca" se refiere a las vacas' es, por cierto, una oración analíticamente relativa a la teoría, pero su pregunta impugnaba *la manera en que se entiende a la teoría*. Dicho enunciado es verdadero en todas las interpretaciones admisibles de la teoría, pero no es eso lo que está en discusión.

El realista interno respondería ahora que (1) se puede argumentar sobre 'la manera de entender la teoría' *dentro* de la teoría, y (2) la cuestión de si esta última posee una *única* interpretación prevista *no* tiene un sentido absoluto. Considerado desde el interior de la Historia 1 (o de un metalenguaje que contenga al lenguaje objeto de la Historia 1), el término 'punto' posee una 'única interpretación prevista'. Considerado desde el interior de la Historia 2 (o de un metalenguaje que contenga al lenguaje objeto de la Historia 2), el término 'punto' *según se emplea en la Historia 1* presenta muchas interpretaciones admisibles. La pregunta '¿cómo lo sabe?' que formuló el crítico presupone un hecho positivo, independiente de la teoría, que se corresponda con un término de la teoría dada; es decir, asume el cuadro del realismo metafísico, y éste es un cuadro que el realista *interno* no necesita (y no debe) aceptar.

El crítico contesta ahora lo siguiente: la 'referencia' ('satisface', hablando estrictamente) es definida de tal modo que:

(1) 'Vaca' se refiere a las vacas

dice solamente (en el caso de que 'vaca' sea una expresión primitiva de *L*) que el par ordenado ⟨'vaca', {vacas}⟩ pertenece a cierta *lista* de pares ordenados. Si acaso, ello *presupone* que 'vaca' posee un referente (bajo *otra* acepción de 'referente'), pero no lo *explica*.[9]

[9] Esta objeción a las definiciones de referencia tarskianas se debe a Hartry Field (*cfr.* su 'Tarski's Theory of Truth', *Journal of Philosophy*, vol. 69, núm. 13, 1972, pp. 347–375) y se expone en mis Conferencias John Locke (primera parte de este libro).

Respuesta: la *utilización* de 'vaca' presupone que esta palabra *se entiende*. Y si mi explicación acerca de la *comprensión* se basara en las *condiciones de verdad* (o *de referencia*), la objeción sería pertinente. Pero se trataba de una explicación verificacionista (según grados de confirmación) y, por lo tanto, *el uso que di al término 'vaca' en el lenguaje* ya ha sido explicado, y estoy en libertad de emplearlo, incluso para dar razón de a qué se refiere 'vaca'.

Lo que estoy diciendo es que, en un sentido 'contextual', el hecho de que 'vaca' se refiera a una determinada clase de cosas (o a una clase de cosas más o menos determinada, pues no tomo en cuenta la vaguedad común del lenguaje) constituye una verdad *a priori*. Adoptar 'lo que se dice de la vaca' equivale a adoptar una 'versión', según la frase de Nelson Goodman, a partir de la cual es una cuestión *a priori* el que la palabra 'vaca' posea un referente (y, en realidad, que se refiere a las vacas).

Uno de los aspectos desconcertantes del cuadro del realista metafísico consiste en que vuelve ininteligible a la posibilidad de que *existan* verdades *a priori*, aun las contextuales, e incluso como un límite (quizás inalcanzable). Una verdad *a priori* tendría que ser producto de una especie de 'intuición' directa de las cosas mismas. Aun la verdad verbal resulta de difícil comprensión. Consideremos el enunciado: 'ninguna persona soltera está casada'. Puede ser 'verbal' el hecho de que, en cierto sentido, esto no resulte lo mismo que 'ninguna persona casada está casada', y, a su vez, ello es un ejemplo de 'todos los *AB* son *A*'. Pero ¿por qué *esto* es verdadero?

Supongamos que no hubiera posibilidad de revisión, y que tal imposibilidad fuera *absoluta*. Y supongamos que sostenemos que la oración 'todos los *AB* son *A*' (e incluso 'todos los solteros son solteros') es completamente inmune a la revisión. ¿Por qué ello la convertiría en *verdadera*?

Aunque sea inimaginable, consideremos que hay algunos *AB* que no son *A*. (Después de todo, existen muchas cosas en la ciencia moderna que resultan inimaginables.) Entonces, bajo el cuadro del realismo metafísico, el negarnos a abandonar la afirmación de que 'todos los *AB* son *A*', no la convierte en verdadera, sólo nos convierte a *nosotros* en obstinados.[10]

[10] El lector puede sentirse tentado a replicar que incluso un metafísico realista tiene derecho a la noción de *convención verbal*. ¿Y por qué no puede ser una *convención* verbal el que uno no ha de referirse a ninguna situación como la *conjunción*

Una vez que abandonamos el cuadro del realista metafísico, la situación se vuelve totalmente diferente. Supongamos que incluimos una oración S dentro de la teoría ideal T_1, sólo porque se trata de una característica que *queremos* para la teoría ideal. (Supongamos que incluso sostenemos que S es 'inmune' a la revisión, como dato conductista con respecto a nosotros.) Considerando que S no hace que T_1 resulte incoherente, T_1 *sigue* teniendo un modelo. Y como el modelo no se encuentra fijado *con independencia* de la teoría, T_1 será *verdadera* —en *el* modelo (desde el punto de vista de *meta-*T_1) y en todos los modelos *admisibles*, desde el punto de vista de una teoría en la cual, para comenzar, los términos de T_1 no poseen un referente determinado. Por lo tanto, S será verdadera. 'S' es 'analítica', pero con una cualidad analítica que recuerda la explicación de Kant sobre lo *sintético a priori* más que su descripción de lo analítico. Porque la oración 'analítica' es, por así decirlo, parte de la 'forma de la representación' y no su 'contenido'. No puede ser falsa con respecto al mundo (como opuesto a EL MUNDO), porque éste no es descriptible con independencia de nuestra descripción.

de las situaciones descritas por las oraciones p, q a menos que impliquen a p y a q por separado? Esto significaría que '$p \cdot q$ implica p' es 'verdadero por convención', y de un modo semejante, 'todos los AB son A' sería 'verdadero por convención'.

La dificultad estriba en que tal 'convención' permite que la conjunción $p \cdot q$ implica p se vuelva analítica sólo *si existe la situación en cuestión*. ¿Pero cómo puede ser *en sí misma* pasible de una *convención*, según el cuadro del realista metafísico, la *existencia* de una situación que posee las propiedades de implicar p e implicar q, además de ser implicada por todas las situaciones que impliquen tanto a p como a q?

Para determinar que lo anterior no es fútil, debo subrayar que existe una lógica (estudiada por David Finkelstein en relación con ciertas teorías físicas 'de avanzada'; *no* la teoría cuántica común) en que (1) hay proposiciones *incompatibles* con cualquier proposición dada, pero (2) no existe nada que se parezca a *la* negación de una proposición dada; es decir, no hay una proposición que sea la más endeble a nivel lógico como para que resulte incompatible con una proposición dada. (Esta lógica corresponde, en términos algebraicos, a redes que no están ortocomplementadas.) Si 'la lógica del mundo' perteneciera a *esta* lógica (como piensa Finkelstein), *la existencia de un complemento* para una situación dada es *falsa como hecho positivo*, y no habría *ninguna* convención lingüística capaz de convertirla en verdadera.

En mi opinión, un realista metafísico coherente o bien debe considerar la lógica como empírica, no sólo en el sentido de creer que puede ser revisada (que es lo que yo pienso), sino también de que no posea *ningún* componente convencional (de tal modo que incluso nuestra convicción de que los enunciados no son *a la vez* verdaderos y falsos se transforma, en última instancia, sólo en una convicción *inductiva*); o bien debe suponer que la lógica es *a priori* en un sentido de *a priori* que no sea explicable de ningún modo por la noción de convención.

Aun si T_1 fuera *incoherente*, si nosotros fuéramos persistentemente incoherentes (por asignar la 'verdad' y la 'falsedad' a las oraciones de una manera homogénea), este argumento no se vería obstaculizado, pues se puede considerar la incoherencia *homogénea* como una *reinterpretación de las conectivas lógicas*. Cuando abandonamos el cuadro realista metafísico vemos, por primera vez, la posibilidad de una verdad 'acerca del mundo' ('todos los *AB* son *A*' *es* 'acerca del mundo', por cuanto trata de todas las *clases A*, *B*) y 'sin contenido'.

En lo que antecede, utilicé como una idealización el concepto de una verdad que carece de toda posibilidad de revisión. Por supuesto, coincido con Quine en que ésta constituye un 'límite' inalcanzable. Cualquier enunciado puede ser 'revisado'. Pero a pesar de que Quine insiste en ello repetidas veces, con frecuencia se pasa por alto que la posibilidad de revisar las leyes de la geometría euclideana o de la lógica clásica no las transforma en simples enunciados 'empíricos'. Ésta es la razón de que yo las haya calificado de *contextualmente a priori*.[11] Quine expresa adecuadamente el problema cuando afirma que 'el saber de nuestros padres' es negro con los hechos y blanco con las convenciones, y añade que no hay hilos *completamente* blancos ni *completamente* negros. Se podría describir lo anterior como un kantismo moderado (y desmitificado). Uno de los problemas del cuadro realista metafísico estriba en que no se entiende cómo es posible que haya algo blanco, ni siquiera blanco grisáceo.

Me permitiré concluir con una última metáfora filosófica. La imagen que Kant tenía del conocimiento era la de una 'representación', una especie de obra teatral. El autor soy yo. Pero el autor también aparece como personaje del drama (igual que en una obra de Pirandello). El autor que aparece en el drama no es el 'real', sino el 'yo empírico'. El autor 'real' es el 'yo trascendental'.

Me gustaría modificar la imagen kantiana de dos maneras. Los autores (en plural, porque mi imagen del conocimiento es social) no sólo escriben *una* historia, sino muchas versiones. Y los autores que aparecen *en* las historias son los *reales*. Esto sería una 'locura' si los relatos fueran *ficciones*. Un personaje ficticio no puede ser también el autor real. Pero se trata de historias verdaderas.

[11] En 'It Aint't Necessarily So', reimpreso en mi *Mind, Language and Reality; Philosophical Papers*, vol. 2, Cambridge University Press, 1975.

ÍNDICE ANALÍTICO

aceptabilidad, 120, 121, 125

aprendizaje, 78, 86, 88, 94, 123, 128, 130; confiabilidad del, 121–123, 125–126, 131

aptitud(es) (habilidades), 77, 89–91

Aristóteles, 13

'aseverabilidad justificada', 9, 42, 49, 61, 119, 127–128

aseverabilidad, 11, 17; justificada, 9, 42, 49, 61, 119, 127–128

Avenarius, 131

Ayer, A.J., 17

Baker, Judy, y Paul Grice, 101–104, 112

Bayes, teorema de, 95

beneficio de la duda, principio del, *véase* caridad, principio de

Berkeley, George, 29

Bohr, N., 34–36

Boyd, Richard, 12, 31–33, 37, 44–45, 50

Brouwer, L. F. J., 37

cálculo proposicional, 38–39

caridad, principio de, 33, 35–37, 52–53, 76, 80

Carnap, Rudolf, 10–11, 17, 78, 115, 118, 131–132, 134, 135

cartesianismo, 102

causa, 13, 26, 53, 68, 76, 87, 123–125, 128–129, 145

causal, explicación, 25–27, 32, 45, 51–52, 126–127, 131; neutralidad filosófica, 126

Celine, L.-F., 108–109

'cerebros en una tina', 143, 146

certeza, 11, 111

Chomsky, chomskianismo, 65–66, 69, 75, 148

ciencia, 12–13, 23–24, 67–69, 89–90, 91–93, 101, 103, 105–106, 107, 110–112, 126–127; como filosofía, 31, 110; desilusión con respecto a la, 105, 106; empírica, 13, 39, 85; éxito de la, 29–30, 32–33, 42, 45; filosofía de la, 31, 34, 50, 56, 85–86, 94; método (metodología) de la, 33, 42, 50, 91–93, *véase también* 'método científico'; términos de la, 32, 33–37, 75, 76, 141, *véase también* teorías de la; y realismo, 31, 50

ciencias morales, 13, 85

ciencias sociales, 13, 77, 86, 105, 110–111; y matemáticas, 78, 96; comparada con la física, 80, 83–84, 86–88,

93–94, 95–97; cientificismo, 90, 96

cientificismo, 13, 31, 90–91, 96

científico(s), 31–32, 106

comprensión, 26, 56, 89, 101, 115–118, 129–130, 135, 137, 142–143, 146, 147–148, 156

computadoras, 82–83, 94, 97, 103, 104

conducta, 54, 62–63, 68, 79, 83, 116–119, 125; éxito de la, 51, 118–121, 127, 130, 133, 137; lingüística, *véase* lingüística, conducta; modelos causales de la, *véase* seres humanos, modelos explicativos; racionalización de la, 26, 54, 58, 63; total, 63, 118–121, 125, 127, 137

conectivas lógicas, 21, 37–44, 47, 49–50, 127, 132, 158

confiabilidad epistémica, *véase* aprendizaje, confiabilidad del

confiabilidad, 52–54, 121–126, 131

confirmación, 10, 18, 51, 86, 94–95, 134, 143, 156

conjunción-negación, traducción de la, 38–39, 41

conocimiento, 13–14, 86, 88–89, 94–96, 110–111, 116, 122, 158; científico, *véase* ciencia; conceptual, 108; implícito, 91–93, 130,; no científico, 92–93, 102, 107, 110; práctico, 13, 89, 91–92, 101–103, 110, 112; público, 90–91; *véase también* epistemología; literatura como fuente del conocimiento; teorías

consenso de los competentes, 90, 97

convencionalidad semántica trivial, 70

convergencia (de las teorías científicas), *véase* teorías, convergencia de las

correspondencia, teoría de la, *véase* verdad

creencia(s), 17, 32, 54, 63–65, 94, 119–122, 125–126, 135–136

Criterio S (de Satisfacción), 43–44, 47

Criterio T (de Verdad), 21–22, 25, 40–41, 43–44, 47, 123, 126–128

cuasi-intuicionismo, *véase* intuicionismo

Dalton, John, 33

definición, 13; explícita, 21; *véase* también verdad, definición de

demostrabilidad, 10, 37–41, 47, 49, 61, 146

descripciones, 35–37, 40–41, 49, 63–65, 67–68, 76, 78– 79, 150, 158; equivalentes, 66–67, 69–71, 150–152; normativas, 56, 62

desentrecomillado, 17–19, 22, 51, 126

Dewey, John, 9

Dilthey, Wilhelm, 93

discusión, 135–137

distinción hecho/valor, 13

Dostoievsky, Feodor, 107–108

Dummett, Michael, 6, 39, 42, 115, 127–129, 146–148

Einstein, Albert, 30, 66, 70

empatía, 86–88, 94–95

empirismo, 9–11, 26, 85–86, 96–97, 108–109; ambiente científico del, 53, 76, 134, 148, representación simbólica del, 12–13, 141; lógico, 49, 85, 90; *véase también* el calificativo empírico bajo hecho; realismo

'en principio', concepto de, 77–78, 81–84, 89, 103

epistémica, confiabilidad, *véase* aprendizaje, confiabilidad del

epistemología, 17–18, 31–32, 50–51, 147;

véase también conocimiento

equilibrio reflexivo, 102

equivalencia, condición de, 10–12

escepticismo, 36, 42, 87–88, 120

espacio, discrecional, 102; explicativo, 57–58, 62

estereotipos, 116, 136

ética, *véase* filosofía moral; ciencias morales

euclideana, geometría, 111, 158

Eurípides, 107

éxito, *véase* conducta; ciencia

explicación, 29–31, 44–45, 54–59, 62–63, 67–68, 78, 83, 85–86; causal, 25–27, 32, 45, 51–52, 126–128, 131; interés relativo de la, 55, 59, 61–62, 67, 69, 73, 126; modelos de, 12, 78–81, 93, 141–142

extensiones, 23, 50, 127, 134–136

falacia idealista, argumento de la, 128–129

fenomenalismo, 49, 120, 131, 133, 146

Feyerabend, P. K., 34–36, 49

Field, Hartry, 5, 12–13, 23–27, 43–45, 51, 76

filosofía moral, 13, 101–105, 110–112

Firth, Roderick, 147

física, 32, 79–80, 92, 110; comparada con las ciencias morales, 13, 85; comparada con las ciencias sociales, 66–67, 69, 77–78, 80, 83, 85–87, 92–93, 95

fisicalismo, 12, 23, 26, 45, 51, 54, 77–79, 82

Freud, freudianismo, 78, 106–107

Garfinkel, Alan, 55–57

gavagai, 53, 57–59, 63–66, 69, 73

Goodman, Nelson, 6–7, 151, 156

Goodman, Paul, 106

Grice, Paul, 65, 101–104, 112

hablante, *véase* lenguaje

Hart, H. L. A. y Honore, A. M., 68

hecho, 132–133, 146–147; empírico, 30, 81–82, 123; positivo, 49, 66, 150, 154–155

Hempel, C. G., 11, 62

Heyting, Arend, 37

historia, 82, 85–88, 92–93, 95–97, 103, 122

Honore, A. M., 68

idealismo, 17, 29, 31, 47, 49–51, 61, 120, 124–125, 127–129, 131

imaginación, 93–94, 96, 103–105, 110, 112

indeterminación, *véase bajo* traducción

inferencia, 24, 39, 87, 89, 91, 94, 124, 130

inobservables, *véase* observables

interés, relatividad del, 55, 59, 61–64, 67–71, 73–74, 126

intuicionismo matemático, *véase* intuicionismo

intuicionismo, 37–40, 41, 44, 50, 61, 146; cuasi-, 40–41, 44, 47, 50, 127

invariantes, 69–70, 150

irrevisabilidad, 155–158

Kant, kantismo, 9, 14, 24, 34, 44, 152, 157–158

Kaufmann, Félix, 10–11

Kripke, Saul, 22, 35, 40, 42, 76, 130

Kuhn, T. S., 34–36, 49

Leeds, Stephen, 24–27, 51

lenguaje, 12–14, 23–24, 89, 91, 137; como una 'mezcla', 117, 136; comprensión del, *véase* comprensión; elaboración de mapas del, 26, 61, 73, 174; formalizado, 10, 18–21, 40, 51- -52; hablantes del, 12–13, 51–52, 54–55, 61–62, 76, 83, 115–116, 126, 141, 148; infinitario, 25; juegos del, 115, 135; - objeto, 10–11, 18–19, 24–26, 40, 117, 130–131, 155; privado, 147; uso del, 13, 115–116, 118, 125–127, 130, 135–137, 141, 147–148, *véase también* significado; y realidad, 26–27, 42–43, 44–45, 118, 123–128, 131, 133, 141–144, 148, 150, 154, *véase también* verdad, teoría de la correspondencia; *véase también* lingüística

Lessing, Doris, 109–110

Lewis, C. I., 132

leyes (de la ciencia, etc.), 30–33, 35, 62–63, 77–80, 82, 85–86, 88, 93, 95, 97

límite ideal, 49, 143

lingüística: competencia, 148; conducta, 53, 118–121, 124- 125, 127, 132, 137; división del trabajo, 116, 134

literatura, 13–14, 101, 103–108, 112; como fuente del conocimiento, 107–110

lógica, 48, 61, 92, 95,; clásica, 37–43, 47, 49, 61, 127, 146, 158; deductiva, 24, 116, 120–121, 126–127, 131–132, 135; inductiva, 20–21, 86–87, 89–90, 115–116, 124, 127, 131–132, 135; no clásica, 37–39

lógico: empirismo, *véase* empirismo

Mach, Ernst, 131

Mailer, Norman, 106

Máquina de Turing (suposición acerca de la), 64, 67

marcos referenciales, teoría de los, 75, 94

Marx, Marxism, 77

matemáticas, 10, 151–152; en las ciencias sociales, 78

McDowell, John, 94, 116

Mendel, Gregor, 33

metafísica, 9–10, 17–18, 64, 95, 110, 112, 119, 121, 127

metalenguaje, 10–11, 18–19, 25, 40, 44, 64, 126, 130, 147, 155; con un número infinito de conjunciones y disyunciones, 24

'método científico' (MC), 85–86, 92–96, 122, *véase también* ciencia, método, metodología

Mill, J.S., 13, 85–86, 88

Minsky, Marvin, 75

Moore, G.E., 128

Murdoch, Iris, 78

Murisaky, Lady, 104

Nagel, Ernest, 85–88, 95

Newton, Isaac, 30, 33

objetividad, 96, 102–103, 111–112, 143–144

observables-inobservables, 42, 131–132, macroobservables, 117

Occam, William, 145

operacionalismo, 29, 32, 75, 109, 143–145

oraciones observacionales, 31, 33, 39, 47–49, 143

organización funcional, 64–71, 78, 81–83

palabras y objetos, *véase* lenguaje y realidad

Peirce, C.S., 9, 49, 143

percepción, 49, 121, 129; visual, 122–123, 128; *véase también* observables

peso, de enunciados (u oraciones), 118, 131–135

Pike, Kenneth, 73–74

platonismo, 85, 95, 111

Polanyi, M., 93

Popper, Karl, 11, 95

posibilidad, 48, 81–84, 95, 127–128, 152

positivismo, 9, 29–31, 93, 120

positivistas del Círculo de Viena, 9

pragmatismo, 10, 56, 62, 120, 145

predicción, 26, 30–33, 39, 42, 56, 81–83, 86, 120

primitiva, noción, referencia, etc., 10, 18–19, 23–25, 40, 44–45; lista de, 19, 21, 25, 43

principio, *véase* concepto 'en principio'

probabilidad, 10–11, 17, 48, 54, 116, 131–132; previa, 95, 124

Protágoras, 42

psicología, 54–57, 61–71, 74, 77–78, 81, 91–95, 103, 105, 107–108, 110–111, 124

puede ser, enunciados sobre lo que, 128

Quine, W.V., 5–7, 17, 26 , 34, 44, 50, 54–55, 57–58, 65–66, 69, 77, 89–90, 152, 158

Rawls, John, 102

realidad, 11; *véase también* hecho; lenguaje y realidad; verdad, dependiente de la mente,

realismo metafísico, 14, 142–143, 145, 147–150, 152, 157– 158

realismo, 41, 66–67, 118, 120, 126–129, 131–134, 137, 148– 150; empírico, 12–13, 30, 141, 148; e idealismo 17, 29, 50–51, 124–125, 127–129, 131; interno, 141, 143, 147, 149, 154–155; metafísico 14, 142–148, 150, 155, 157– 158; y ciencia, 49–50; 'científico', 30–32; y verdad, 9, 11–12, 29–32, 42, 44, 47, 49–50, 61, 117, 121–123, 146–147

referencia, 14, 20–24, 26, 33–36, 42–45, 49–53, 61, 75–76, 115, 117–118, 126, 130–131, 134–136, 142–143, 145, 147–148, 153–156; en cuanto una noción causal explicativa, 25, 27, 32, 45; indeterminación, 58, 61–62, 93; teoría fisicalista, 13, 45, 51; primitiva, *véase* primitiva

Reichenbach, Hans, 115, 118, 131, 133–135

relatividad explicativa, 57

relatividad ontológica, 152

Russell B., 76, 132

Schiffer, R.S., 65

Sellars, W.S., 49

semántica, 10–11, 18, 23, 33, 42, 61, 124, 127, 130, 133, 142, 146–147; no realista/verificacionista, 146–148; referencial, 32–33, 35

seres humanos, 42, 82, 89, 97, 107; ataque de los humanistas a los científicos, 105–106; conducta de los, *véase* conducta; modelos explicativos de los, 12, 78–81, 93; organización funcional de los, *véase* organización funcional

significado, 18, 50, 52, 53, 54, 56, 58, 62, 88–90, 118, 130, 132–133, 135–137, 148; cambio de, 117, 134–135; 'mito del museo', 65; teoría pictórica, 115, 118; teoría probabilística, 131–134; y realismo, 132–135; y uso, 11, 115, 117, 132, 137; *véase también* uso del lenguaje

'significados motivadores', 58, 89

signos, 145–146

sociología, 77, 92–93, 95–97

Spinoza, Benedicto, 111

Sutton, Willie, 42, 57, 62

Tarski, tarskismo, 5, 9–10, 12–13, 17–23, 25–26, 40–45, 47, 52, 61, 123, 127, 130, 141, 147

teoría de la relatividad, 30, 33, 70, 111

teorías, 30, 34, 41–42, 66–67, 79, 91, 96, 120, 123–125, 143, 151–155, 157; candidatas, 32, 34; convergencia de, 30–31, 42, 49–50, 141; ideal, 49, 143–145, 149, 154, 156–157; inconmensurabilidad, 35; total, 50–51; sopesamiento de, 94; *véase también* teoría de la relatividad

tiempo real, 82–83

traducción, 26, 45, 51–52, 54–55, 57–59, 61, 63, 65, 66, 70– 71, 74–78, 88, 91–92, 94, 115, 153–154; de la conjunción/negación, 38–39, 41; indeterminación de la, 55, 58–59, 64, 69, 73–74, 90; mecánica, 75, 77, 89, 94; radical, 57, 73–75, 119, 126

Tynan, Kenneth, 106

utilidad, 51, 116, 131

utopismo científico, 76–77, 95

valencia, definición de, 23, 25, 45

verdad, 10–14, 17, 24–25, 37, 40–45, 48, 50–52, 61, 80, 92, 115, 117–118, 120–121, 123–128, 130, 132, 137, 143, 158; a priori, 122, 156, 158; aproximada, 10–11, 30–33, 80; condiciones de 53–54, 57, 118, 129–130, 133, 135, 137, 142, 146–148, 156; teoría de la correspondencia, 9, 11–13, 29–30, 41–43, 115, 117, 129, *véase también* lenguaje y realidad; definición, 10, 19–21, 25, 41–42, 45, 47, 50, 52–54, 123–126; funciones, 19–20, 25, 39, ; dependiente de la mente, 9, 14; neutralidad filosófica, 11–12, 17–18, 50; 'concepción realista' *véase* realismo; valores de verdad, 26, 37, 127, 151; 'concepción verificacionista' *véase* verificacionismo

verificación, 10–11, 18, 62, 91, 117, 128, 136, 143, 146– 148

verificacionismo, 9, 13, 131–134, 156, 147–149

verosimilitud, 95, 108–109, 144

Verstehen, 86–87, 93–95

Weber, Max, 93

Winch, Peter, 62, 64

Wittgenstein, Ludwig, 115, 118, 130, 136, 147

Zilsel, Edgar, 85

ÍNDICE

Prefacio 5

Introducción 9

Primera parte. Significado y conocimiento (conferencias John Locke, 1976) 15

- Primera conferencia 17
- Segunda conferencia 29
- Tercera conferencia 47
- Cuarta conferencia 61
- Quinta conferencia 73
- Sexta conferencia 85

Segunda parte. Literatura, ciencia y reflexión 99

Tercera parte. Referencia y comprensión 113

Cuarta parte. Realismo y razón 139

Índice analítico 159

El significado y las ciencias morales se terminó de imprimir el 15 de diciembre de 1991 en los talleres de Regina de los Ángeles, S. A. Para su composición y formación, realizadas por computadora en el Instituto de Investigaciones Filosóficas, se utilizó el programa TEX y tipos Dutch de Bitstream. La edición estuvo al cuidado de Francisco Hernández Quiroz y consta de 2 000 mil ejemplares.

El profesor Putnam ha impulsado una concepción del conocimiento que considera que la ética, el conocimiento práctico y las partes no-matemáticas de las ciencias sociales (como la historia) son partes del conocimiento en la misma medida que la física. A su vez, rechaza la idea de que el conocimiento pueda ser separado del no-conocimiento por el hecho de que sólo el primero pertenece a algo llamado "el método científico". Toda la discusión se ubica en el contexto de una explicación de la verdad que se parece a la de Kant, ya que insiste tanto en el realismo empírico como en la dependencia que guarda todo conocimiento con respecto a la mente.

La primera parte del libro contiene las conferencias que el profesor Putnam dictó en la cátedra John Locke de la Universidad de Oxford, en 1976. Éstas ofrecen el estudio detallado de una teoría "fisicalista" sobre la referencia, en el contexto de las obras de Tarski, Carnap, Popper, Hempel y Kant. Sin embargo, el análisis va más allá, hasta los conceptos de verdad, y problemas de carácter lingüístico y de ciencias sociales en general. Los artículos que completan el volumen cubren una serie de temas diferentes pero que tienen que ver con el gran problema metafísico del realismo, de la naturaleza del lenguaje y la referencia, y del carácter de nosotros mismos, temas que se traslapan, se relacionan entre sí e intercambian información.

SIGNIFICADO CIENCIAS
9 789683 622501